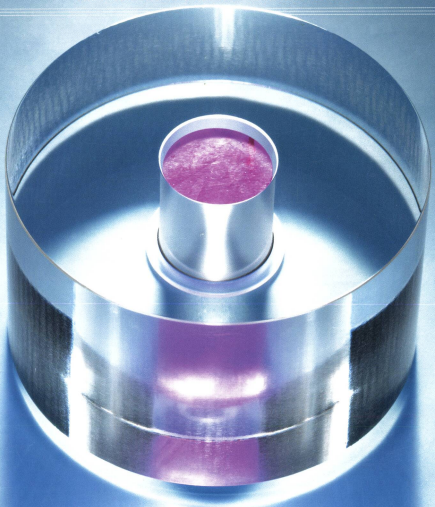
INSIGHT
Chemistry and energy

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



RETURN OF THE MASER

Bench-top 'radio laser' seeks practical applications PAGES 285 & 353

GENETICS

SCIENCE ON THE SILK ROAD

Seeking the genetics of food preference

PAGE 269

TECHNOLOGY

BUILDINGS FOR THE FUTURE

Time to invest in energy-efficient construction

PAGE 275

CANCER THERAPY

SECOND SITE

'Passenger' mutations as surrogate drug targets

PAGES 284 & 337

NATURE.COM/NATURE
16 August 2012 £10
Vol. 488, No. 7411

SCIENTIFIC REPORTS

Report Card

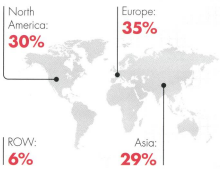
In the first year of publication*

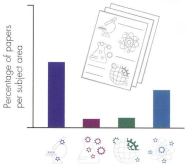
SPEED

99 days on average from submission to publication

DIVERSITY

457 papers published in the first year



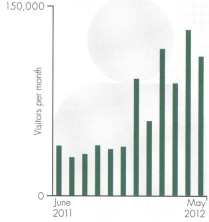


VISIBILITY

848,000 total visitors

1.89 million total page views

3092 average page views per paper



COMMUNITY

Facebook likes: **32,000+**

93% of authors would submit again†

Keep your research moving. Submit to *Scientific Reports*!

nature.com/scientificreports

nature publishing group npg

*All data as of June 2012. †Author survey.

I M P A C T F A C T O R

NPG Asia Materials has received its first impact factor. In a short time, *NPG Asia Materials* has established itself as an essential resource for the materials science community!

Good reasons to submit your next excellent paper to *NPG Asia Materials*:

- Ranked 20 of 231 in Multidisciplinary Materials Science*
- Open Access – anyone can download and read your paper
- High exposure on nature.com
- Widely read by the global audience

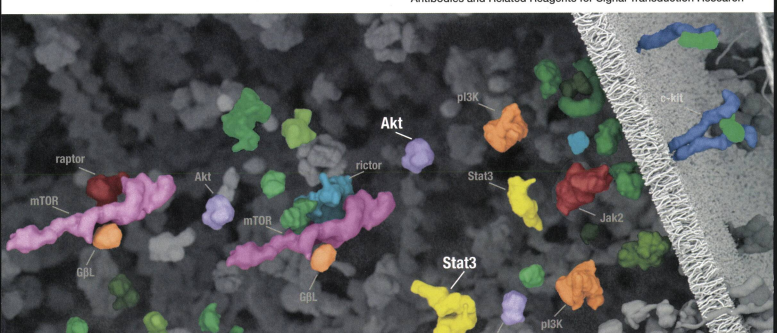
OPEN ACCESS to all articles published in *NPG Asia Materials* – free and published under Creative Commons license

N A T U R E . C O M / A M

npg asia materials
founded in association with the Tokyo Institute of Technology

nature publishing group npg

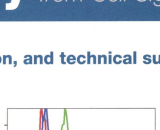
* 2011 Journal Citation Reports® (Thomson Reuters, 2012)



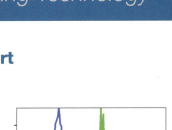
Conjugated Antibodies for Flow Cytometry *from Cell Signaling Technology*

Unparalleled product quality, validation, and technical support

- The highest quality antibodies provide you with the brightest signal and the lowest background.
- Extensive validation by our in-house flow cytometry group means optimization is not left up to you, the user.
- Technical support provided by our flow cytometry specialists translates into a thorough, fast, and accurate response.
- High quality custom conjugations of our off-the-shelf antibodies are also available, please inquire about sizing and pricing options.
- Multiple fluorochrome options are available to meet your unique multiplexing needs.



Phospho-Stat3 (Tyr705) (D3A7) XP® Rabbit mAb (Alexa Fluor® 488 Conjugate) #4323: Flow cytometric analysis of Jurkat cells, untreated (blue) or IFN-α treated (green), using #4323 compared to nonspecific control antibody (red).



Phospho-Akt (Ser473) (D9E) XP® Rabbit mAb (PE Conjugate) #5315: Flow cytometric analysis of Jurkat cells, untreated (green) or treated with LY294002 (PI3 Kinase Inhibitor) #9901, Wortmannin (PI3 Kinase Inhibitor) #9951, and U0126 (MEK1/2 Inhibitor) #9903 (blue), using #5315.

©2010 Cell Signaling Technology, Inc. | XP® and Cell Signaling Technology® are trademarks of Cell Signaling Technology, Inc. Alexa Fluor® is a trademark of Molecular Probes, Inc.

for quality products you can trust...

www.cellsignal.com

Cell Signaling TECHNOLOGY®

Orders (toll-free) 1-877-616-2355 | **Technical support** (toll-free) 1-877-678-8324 support@cellsignal.com | **Inquiries** info@cellsignal.com | **Environmental Commitment** eco.cellsignal.com

For Research Use Only. Not For Use In Diagnostic Procedures.

CONTENTS

16 August 2012 / Vol 488 / Issue No 7411

THIS WEEK

EDITORIALS

253 SCIENCE & SOCIETY
Repeat after me
A bid by Romanian academics to quash plagiarism deserves support

253 HEALTH & SAFETY
Safety shambles
The dirty truth about the Fukushima clean-up

WORLD VIEW

255 **Why we are poles apart on climate change**
Dan Kahan
Pollution in the communication of climate science drives cultural divisions

RESEARCH HIGHLIGHTS

256 SELECTIONS FROM THE SCIENTIFIC LITERATURE

Hox-linked aggression / Hyena curiosity / Picture perfect / Whole-cell model / Sun-seeking solar cells

SEVEN DAYS

258 THE NEWS IN BRIEF
Hottest month on record for the US / BP backs materials research / Indian politicians pour cold water on GM crops / Elan spins off drug-discovery arm / US cleans up dioxins at site in Vietnam / Indiana sees outbreak of swine flu

CAREERS

419 TRAINING
Workshops that work
It pays to make the most of any skills training that comes your way

421 TURNING POINT
Laskar scholar Nehal Mehta balances clinical and research aspects of cardiology

421 CAREER BRIEFS

NATUREJOBS ADVERTISING FEATURE
Spotlight on chemistry

NEWS IN FOCUS

261 ENVIRONMENT
Mass protests power green policies in China

262 PLANETARY SCIENCE
Curiosity's chief scientist sets out his vision for Mars rover

264 MISCONDUCT
Independent committee aims to tackle scientific misconduct in Romania

265 INFECTIOUS DISEASES
Ugandan lab proves its worth in Ebola outbreak

266 CHEMISTRY
Amino-acid catalyst speeds synthesis of blockbuster drug family

267 AIDS
Patent expiries promise to make HIV treatments more affordable

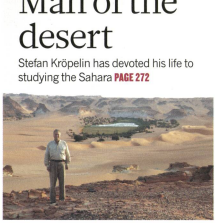
FEATURES

269 FOOD
Taste for adventure
A trip along the Silk Road reveals how people's love of food evolved

GEOLOGY

Man of the desert

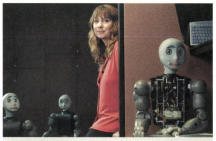
Stefan Kröpelin has devoted his life to studying the Sahara PAGE 272



COMMENT

275 ENGINEERING
How to build a low-energy future
Philip Farese
Strategic investment in emerging technologies could save vast amounts of energy — and cash

BOOKS & ARTS



Q&A

Helping hands

Social roboticist Maja Matarić on robots that care PAGE 280

278 COSMOLOGY
Enigma at the heart of the Universe
Mario Livio

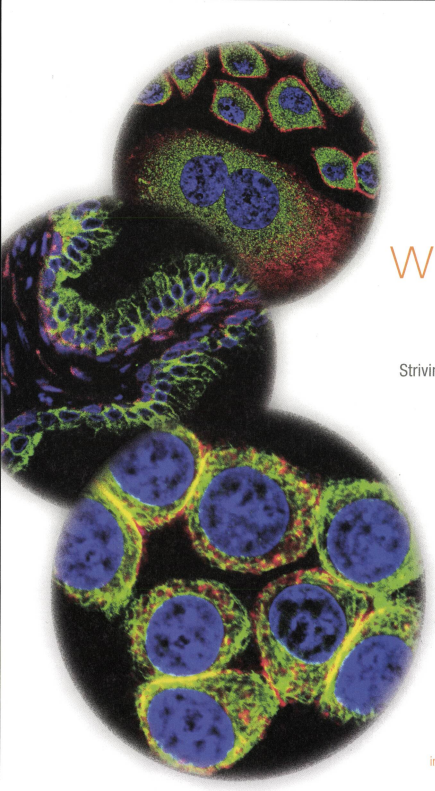
279 TECHNOLOGY
Pre-digital dreams
Josie Glausiusz

CORRESPONDENCE

281 Political meddling threatens Israeli universities / Talking Turkey / The deceased author problem / The truth about nanomaterials

FUTURES

424 **Tige is the man**
Sue Lange



Proteome-wide antibody coverage.

Striving towards an antibody for each human protein.

Unsurpassed development capabilities built over the last ten years allow Abgent to produce more than 5,000 new antibodies per year. Protein A and peptide affinity purification, along with rigorous antibody validation, ensures antibody quality and assay flexibility. The end result is a broad offering of in-house developed, high quality antibodies tailored to your research and drug discovery needs.

Applications: Western Blot, Immunohistochemistry, Immunofluorescence, Flow Cytometry, ELISA, and more.

20µg Evaluation Size Vials
available for most new products.

Visit our website at **www.abgent.com** for more information on Abgent's product and service offerings.

© 2012 ABGENT. All rights reserved. It's all about New, is a trademark of ABGENT. For Research Use Only. Not For Use In Diagnostic Procedures. ABG-00104-0612

It's all about New.™
Unsurpassed Development Capabilities

ABGENT ANTIBODIES ARE ESSENTIAL TOOLS FOR THE STUDY OF CANCER, NEURODEGENERATIVE DISEASES, METABOLIC SYNDROME, IMMUNOLOGY, APOPTOSIS, AUTOPHAGY, DEVELOPMENT AND MORE.

Visit our website for a complete product listing.

www.abgent.com

药明康德 WuXi AppTec | ABGENT

CONTENTS

16 August 2012 / Vol 488 / Issue No 7411

RESEARCH

NEW ONLINE

283 Papers published this week at nature.com

NEWS & VIEWS

284 CANCER
Exploiting collateral damage
Passenger mutations expose cellular vulnerabilities for drug targeting
Ben Lehner & Solip Park
SEE ARTICLE P. 337

285 APPLIED PHYSICS
Masers made easy
A solid-state maser that operates at room temperature
Aharon Blank SEE LETTER P. 353

286 CELL BIOLOGY
Neither brown nor white
A new type of heat-generating fat cell in mice and humans
Barbara Cannon & Jan Nedergaard

288 ASTROPHYSICS
One of the first of the second stars
A primordial star in the constellation Leo
John Cowan

289 NEUROSCIENCE
Circuits drive cell diversity
Heterogeneity within neuron types is partly due to network affiliation
Nathaniel Urban & Shreejoy Tripathy
SEE LETTER P. 375

290 OCEANOGRAPHY
The trouble with the bubble
Revised measurements of rates of marine nitrogen fixation
Angelicque E White SEE LETTER P. 361

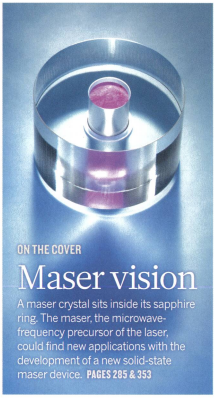
ARTICLES

337 CANCER **Passenger deletions generate therapeutic vulnerabilities in cancer**
F L Muller et al. SEE N&V P. 284

343 NEUROSCIENCE **Division and subtraction by distinct cortical inhibitory networks *in vivo***
N R Wilson, C A Runyan, F L Wang & M Sur

LETTERS

349 ASTRONOMY **A massive, cooling-flow-induced starburst in the core of a luminous cluster of galaxies**
M McDonald et al.



ON THE COVER

Maser vision

A maser crystal sits inside its sapphire ring. The maser, the microwave-frequency precursor of the laser, could find new applications with the development of a new solid-state maser device. PAGES 285 & 353

353 APPLIED PHYSICS **Room-temperature solid-state maser**
M Oxborrow, J D Breeze & N M Alford SEE N&V P. 285

357 MATERIALS **Electronic read-out of a single nuclear spin using a molecular spin transistor**
R Vincent, S Klyatskaya, M Ruben, W Wernsdorfer & F Balestro

361 OCEANOGRAPHY **Doubling of marine dinitrogen-fixation rates based on direct measurements**
T Großkopf et al. SEE N&V P. 290

365 CLIMATE **More extreme swings of the South Pacific convergence zone due to greenhouse warming**
W Cai et al.

370 ANTHROPOLOGY **Reconstructing Native American population history**
D Reich et al.

375 NEUROSCIENCE **A biophysical signature of network affiliation and sensory processing in mitral cells**
K Angelo et al. SEE N&V P. 289

379 NEUROSCIENCE **Activation of specific interneurons improves V1 feature selectivity and visual perception**
S-H Lee et al.

384 IMMUNOLOGY **Bacterial virulence proteins as tools to rewire kinase pathways in yeast and immune cells**
P Wei et al.

389 IMMUNOLOGY **NLRP6 negatively regulates innate immunity and host defence against bacterial pathogens**
P K Anand et al.

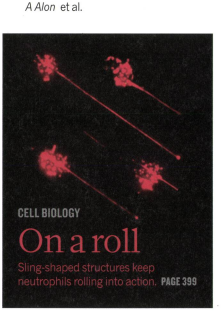
394 DISEASE **APJ acts as a dual receptor in cardiac hypertrophy**
M C Scimia et al.

399 CELL BIOLOGY **'Slings' enable neutrophil rolling at high shear**
P Sundd et al.

404 STEM CELLS **A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response**
L Kruidenier et al.

409 STEM CELLS **The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming**
A A Mansour et al.

414 STRUCTURAL BIOLOGY **The dynamic disulphide relay of quiescin sulphydryl oxidase**
A Alon et al.



CELL BIOLOGY

On a roll

Sling-shaped structures keep neutrophils rolling into action. PAGE 399

SCIENTIFIC AMERICAN™

Gain complete online access to the latest in science, technology and medicine at your institution

Recommend site license access to your library today!
www.nature.com/scientificamerican

™ Scientific American, Inc.

nature publishing group npg

INSIGHT

16 August 2012 / Vol 488 / Issue No 7411

CHEMISTRY AND ENERGY

PERSPECTIVE

294 **Opportunities and challenges for a sustainable energy future**
Despite the significant growth in the use of renewable energy in the past two decades, there has been little change in the amount of oil, coal and natural gas we consume annually. Many of the alternatives to carbon-emitting sources are still in the research and development stage, but these research areas may provide the answer to our future energy needs. Finding ways for these alternative forms of energy to fit into the existing energy landscape in a cost-effective way will be important in securing a sustainable energy future.
Steven Chu & Arun Majumdar

REVIEWS

304 **Materials interface engineering for solution-processed photovoltaics**
The rapid progress made in solution-processed photovoltaics has increased the efficiency of some of the devices to more than 10%. The interfaces present within the variety of different solution-processed technologies have a crucial effect on device performance. A better understanding of the processes that occur at the interfaces and their structural control through materials chemistry will be essential to improve performance, and could allow photovoltaics to become a cost-efficient way of harnessing the solar power that irradiates Earth.
Michael Graetzel et al.

313 **Membrane-based processes for sustainable power generation using water**
Water is essential for many processes used to generate power, but using the water itself to produce the energy could provide a source of sustainable and renewable power. From the various systems being developed, three promising methods stand out: pressure-retarded osmosis, reverse electrodialysis and microbial fuel cells. These methods could one day become viable and cost-effective ways to capture energy from sea water and waste water.
Bruce E. Logan & Menachem Elimelech

320 **Microbial engineering for the production of advanced biofuels**
Biofuels produced by microorganisms could 'drop in' to the existing transport infrastructure, but only if these organisms could be engineered to produce fuels in yields high enough to compete with petroleum-based fuels. Overcoming the endogenous self-regulation of pathways that produce the biofuel in native hosts and reconstructing the pathways in a heterologous host are the main challenges that need to be overcome. A better understanding of the enzymes and metabolic pathways involved in the production of biofuels should allow engineered microorganisms to compete commercially with more-conventional fuels.
Pamela P. Peralta-Yahya et al.

329 **Exploiting diversity and synthetic biology for the production of algal biofuels**
Algae can produce oils using only sunlight, carbon dioxide and water. This natural ability can be harnessed to produce biofuels as an alternative to fossil fuels. Finding the ideal algal strain is made easier as a result of the group's huge diversity; however, finding a natural strain that meets the many requirements for efficient large-scale biofuel production is unlikely. In a similar way to their use in improving plant crops, breeding programmes and genetic engineering could be used to develop strains of algae that are suitable for agricultural and industrial cultivation.
D. Ryan Georgianna & Stephen P. Mayfield

nature

16 August 2012 / Vol 488 / Issue No. 7411

EDITORIAL

LONDON, MUNICH, WASHINGTON DC, NEW YORK, BOSTON, PARIS, SAN FRANCISCO, TOKYO nature@nature.com

EDITOR-IN-CHIEF: Philip Campbell
EXECUTIVE EDITOR: Nick Campbell
PUBLISHING EXECUTIVE EDITOR: Maxine Clarke **EXECUTIVE EDITOR:** Véronique Kiermer
CHIEF MAGAZINE EDITOR: Tim Appenzeller
THIS WEEK/NEWS/FEATURES: Alison Abbott, David Adam, Ananyo Bhattacharya, Geoff Brumfiel, Declan Butler, Ewen Callaway, Erika Check Hayden, Daniel Cressey, David Cyranoski, Natasha Gilbert, Eric Hand, Heidi Ledford, Corie Lok, Brendan Maher, Emma Marris, Richard Monastersky, Brian Owens, Helen Pearson, Mark Peplow, Eugenie Samuel Reich, Ivan Semeniuk, Quirin Schiermeier, Jeff Tollefson, Richard Van Noorden, Meredith Wadman, M Mitchell Waldrop **COMMENT:** Sara Abdulla, Joanne Baker, Rosalind Cotter, Barbara Kiser, Alison McCook, Lucy Odling-Smee, Sarah Tomlin **NATURE PODCAST/VIDEO:** Adam Rutherford, Thea Cunningham, Geoff Marsh, Kerri Smith, Charlotte Stoddart **NEWS AND VIEWS:** Sadaf Shadan, Ana Lopes, Andrew Mitchinson, Cesar Sanchez, Marian Turner **RESEARCH:** Ritu Dhand, Karl Ziemelis, Francesca Cesari, I-han Chou, Tanguy Chouard, Rosamund Daw, Alex Eccleston, Angela Eggleston, Joshua Finkelstein, Henry Gee, Patrick Goymer, Noah Gray, Marie-Thérèse Heemeli, Magdalena Helmer, Karen Howell, Claudia Lupp, Barbara Marte, Andrew Mitchinson, Juliane Mössinger, Deepa Nath, Wayne Peng, Leslie Sage, Magdalena Skipper, Clare Thomas, Maria Trajkovska, John VanDecar, Liesbeth Venema, Ursula Weiss, Michael White **INSIGHTS/REVIEWS/PERSPECTIVES:** Ursula Weiss **CAREERS:** Gene Russo, Karen Kaplan **TECHNOLOGY FEATURES:** Monya Baker **FUTURES:** Colin Sullivan **SUPPLEMENTS:** Herb Brody, Michelle Grayson, Tony Scully **SUBEDITORS:** Colin Sullivan, Sarah Archibald, Nicola Bailey, Anne Blewett, Nicolas Fanget, Isobel Flanagan, Paul Fletcher, Sasha Lewis, Dinah Loon, Louisa Lyon, Anna Novitzky, Yukie Ozawa, Katherine Perry, David Price, Jenny Rooke, David Winters, Anna York **PRODUCTION:** Jenny Henderson, Leonora Dawson-Bowling, Dan Foley, Stewart Fraser, Susan Gray, Simon Gribbin, Hazel Mayhew, James McQuat, Emilia Oviss, Yvonne Strong, Robert Sullivan, Roberta Tarran, Charles Wenz **ART AND DESIGN:** Kelly Buckheit Krause, Wesley Fernandes, Madeline Hutchinson, Barbara Izdebska, Paul Jackman, Jasiek Krzysztofiak, Annthea Lewis, Chris Ryan, Nik Spencer, Claire Welsh, Mark Wood **JAPAN EDITION:** Eiji Matsuda, Mika Ishida, Mitsuyo Utsugi, Sayaka Ayame **PRESS OFFICE:** Ruth Francis, Neda Afsarmanesh, Lisa Boucher, Rachel Twinn, Rebecca Walton **ADMINISTRATION:** Paul Sng, Liz Bankes, Roseann Campbell, Peter Cary, Fion Cleverley, Dana Coles, Faye Fornasier, Ashlee Gaqui, Amy Invernizzi, Caroline McLean, Rina Nozawa, Anastasia Panoutsou, Kenneth Simpson, Diane Yorke

PUBLISHING

LONDON feedback@nature.com

MANAGING DIRECTOR & PUBLISHER: Steven Inchcoombe
ASSOCIATE PUBLISHER: Arash Hejazi
PUBLISHING PROJECT MANAGERS: Claudia Deasy, Christian Manco

TOKYO feedback@nature.asia

REGIONAL MANAGING DIRECTOR — ASIA-PACIFIC: David Swinbanks
DIRECTOR — ASIA-PACIFIC: Antoine E. Bocquet
OPERATIONS DIRECTOR — ASIA-PACIFIC: Hiroshi Minemura

DISPLAY ADVERTISING, SPONSORSHIP & NATUREJOBS

GLOBAL HEAD: Andrew Douglas a.douglas@nature.com

DISPLAY ADVERTISING AND SPONSORSHIP
GLOBAL HEAD: Gerard Preston g.preston@nature.com
ASSISTANT SPONSORSHIP MANAGER: Reya Silao Tel: +44 (0)20 7843 4977 (rsilao@nature.com)

EUROPE/REST OF WORLD display@nature.com Tel: +44 (0)20 7843 4960
GERMANY/SWITZERLAND/AUSTRIA: Sabine Hugi-Furst Tel: +41 52761 3386
UK/IRELAND/SCANDINAVIA/SPAIN/PORTUGAL: Evelina Rubio Hakansson Tel: +44 (0)20 7843 4079
FRANCE/BELGIUM/THE NETHERLANDS/LUXEMBOURG/ITALY/OTHER EUROPE: Nicola Wright Tel: +44 (0)20 7843 4959

NORTH AMERICA display@natureny.com
NEW ENGLAND/ EASTERN CANADA: Sheila Reardon Tel: +1 617 494 4900
NEW YORK/ MID-ATLANTIC/SOUTHEAST: Jim Breault Tel: +1 212 726 9334
MIDWEST: Mike Rossi Tel: +1 212 726 9255
WEST COAST/ WESTERN CANADA: George Lui Tel: +1 415 781 3804
WEST COAST NORTH: Kayla McCutchan Tel: +1 415 403 9038

ASIA-PACIFIC display@nature.asia
ASIA-PACIFIC HEAD: Kate Yoneyama
JAPAN: Ken Mikami Tel: +81 3 3267 8765
GREATER CHINA/SINGAPORE: Gloria To Tel: +852 2811 7191
INDIA: Vikas Chawla Tel: +91 124 288 1057

SPONSORSHIP
BUSINESS DEVELOPMENT MANAGER: David Bagshaw Tel: +1 212 726 9215
BUSINESS DEVELOPMENT MANAGER: Patrick Murphy Tel: +1 617 475 9216

NATUREJOBS naturejobs@nature.com
EUROPEAN HEAD OFFICE, LONDON: Tel: +44 (0)20 7843 4961
EUROPEAN SALES MANAGER: Nils Moeller (n.moeller@nature.com) **BUSINESS DEVELOPMENT MANAGER, MIDDLE EAST:** Jon Giuliani (j.giuliani@nature.com) **ADVERTISING PRODUCTION MANAGER:** Stephen Russell **NATUREJOBS ONLINE PRODUCTION:** Dennis Chu
US HEAD OFFICE, NEW YORK: naturejobs@natureny.com **US SALES MANAGER:** Ken Finnegan
JAPAN HEAD OFFICE, TOKYO: Tel: +81 3 3267 8765 **ASIA-PACIFIC HEAD:** Kate Yoneyama
ASIA-PACIFIC SALES & BUSINESS DEVELOPMENT MANAGER: Yuki Fujiwara **BUSINESS DEVELOPMENT MANAGER, GREATER CHINA, SINGAPORE:** Gloria To Tel: +852 2811 7191
INDIA: Vikas Chawla Tel: +91 1242881057 (v.chawla@nature.com)

MARKETING & SUBSCRIPTIONS

USA/CANADA/LATIN AMERICA subscriptions@us.nature.com
Tel: (USA/Canada)+1 866 363 7860; (outside USA/Canada) +1 212 726 9223
MARKETING: Sara Girard **FULFILMENT:** Karen Dawson

JAPAN/CHINA/KOREA subscriptions@natureasia.com
Tel: +81 3 3267 8751 **MARKETING:** Sachiyo Ikeda **FULFILMENT:** Masayo Tanaka

EUROPE/REST OF WORLD subscriptions@nature.com
Tel: +44 (0)1256 329242
MARKETING: Elena Woodstock, Marcus Parker, Hannah Phipps **FULFILMENT:** Jean Burrows

INDIA npgindia@nature.com
Tel: +91 124 2881053/54
HEAD OF BUSINESS DEVELOPMENT, INDIA: Debashish Brahmachari **MARKETING:** Harpal Singh Gill

ANNUAL SUBSCRIPTIONS (UK/Europe, including post and packing)
FULL RATE: £1,400
PERSONAL RATES: UK £135; Europe €209; Rest of World £210
STUDENT RATES: UK £79; Europe €122; Rest of World £125
Individual rates apply to payments by personal cheque or credit card. Back issues: £20.00.

ANNUAL SUBSCRIPTIONS (USA, including post and packing)
INSTITUTIONAL/CORPORATE RATE: $2,730 ; **PERSONAL RATE:** $199; **STUDENT RATE:** $99;
POSTDOC RATE: $119
Printed in USA. Individual rates available only to subscribers paying by personal check or credit card. Orders for student/postdoc subscriptions must be accompanied by a copy of student ID. Rates apply to USA, Canada, Mexico/Central & South America. Add 7% GST tax in Canada (Canadian GST number 140911595). **BACK ISSUES:** US$20.00.

ANNUAL SUBSCRIPTIONS (Asia, including post and packing)
JAPAN: 本間購読料金: *Nature* のみ ¥37,800 円 +21,890 円 (消費税) =459,690 円 (合計)
Nature + *Nature Digest* 466,550 円 +23,328 円 (消費税) =489,878 (合計)
お問い合わせ 日本出版貿易株式会社 国内営業 1 部
〒101-0064 東京都千代田区猿楽町 1-2-1 Tel: 03-3292-3767 Fax: 03-3292-8766
個人 / 学生購読について
NPG ネイチャー アジア・パシフィック フルフィルメント & カスタマーサービス
〒162-0843 東京都新宿区市谷田町 2-37 千代田ビル Tel: 03-3267-8759 Fax: 03-3267-8752
CHINA: www.natureasia.com/cn/subscribe; **KOREA:** www.natureasia.com/korea/subscribe

NATURE ONLINE WWW.NATURE.COM/NATURE

CHIEF TECHNOLOGY OFFICER: Howard Ratner **ASSOCIATE DIRECTOR, NATURE.COM:** Daniel Pollock
PLATFORM DEVELOPMENT: Jeremy Macdonald **WEB PRODUCTION:** Alexander Thurrell, Deborah Anthony **WEB DESIGN:** Juan-Carlos Sobrino, Luke Stavenhagen **PLATFORM INTERFACES:** Amanda Ward

SITE LICENCES, FULFILMENT & CUSTOMER SERVICES

SITE LICENCES: npg.nature.com/libraries feedback@nature.com
CHIEF OPERATIONS OFFICER: John Carroll **ASSOCIATE DIRECTOR OF OPERATIONS:** Dominic Pettit
CUSTOMER SERVICE: Liz Davidson **ORDER PROCESSING OPERATIONS:** Steve Poulter

NATURE PUBLISHING GROUP OFFICES

LONDON: The Macmillan Building, 4 Crinan Street, London N1 9XW Tel: +44 (0)20 7833 4000 Fax: +44 (0)20 7843 4596/7

BASINGSTOKE: Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants RG21 6XS, UK Tel: +44 (0)1256 329242

WASHINGTON DC: 968 National Press Building, 529 14th St NW, Washington DC 20045-1938 Tel: +1 202 737 2355

NEW YORK: 75 Varick St, 9th Floor, New York, NY 10013-1917 Tel: +1 212 726 9200

BOSTON: 25 First Street, Suite 104, Cambridge, MA 02141 Tel: +1 617 475 9275

SAN FRANCISCO: 225 Bush Street, Suite 1453, San Francisco, CA 94104 Tel: +1 415 403 9027

MUNICH: Nymphenburger Strasse 14, D-80335 München Tel: +49 89 549057-13

PARIS: Moulin de Malicorne, Sainte Maure de Touraine, Paris 37800 Tel: +33 2 47 35 72 15

TOKYO: Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843 Tel: +81 3 3267 8751

INDIA: Nature Publishing Group, 3A, 4th Floor, DLF Corporate Park, Gurgaon 122002 Tel: +91 124 2881053/54

Nature® (ISSN 0028-0836) is published by Nature Publishing Group, a division of Macmillan Publishers Ltd (The Macmillan Building, 4 Crinan Street, London N1 9XW). Registered as a newspaper at the British Post Office. North and South American orders to subscriptions@natureny.com (Tel: +1 866 363 7860). Other orders to subscriptions@nature.com (Tel. +44 (0)1256 329242; Fax +44 (0)1256 812358). Authorization to photocopy material for internal or personal use, or internal or personal use of specific clients, is granted by *Nature* to libraries and others registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided the relevant copyright fee is paid direct to CCC, 222 Rosewood Drive, Danvers MA 01923, USA. Identification code for *Nature*: 0028-0836/03. CPC PUB AGREEMENT #40032744. In the US, *Nature* (ISSN 0028-0836) is published weekly on Thursday, except the last week in December by Nature Publishing Group, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. US Periodicals postage paid at New York NY, and additional mailing post offices. US POSTMASTER: send address changes to *Nature*, Subscription Dept, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. Published in Japan by NPG Nature Asia-Pacific, Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843, Japan. © 2012 Macmillan Publishers Limited. All rights reserved.

THIS WEEK

EDITORIALS

WORLD VIEW Why climate sceptics are rational yet wrong **p.255**

PICTURE PERFECT Nanotechnology brings out the best in colour images **p.256**

LIFE AQUATIC Bubbly feet stop leaf beetles from sinking fast **p.257**

Repeat after me

With plagiarism seemingly endemic in Romania, as well as rife among Europe's political class, a bid by academics to root out misconduct deserves widespread support.

Elena Ceaușescu did not have a BSc, but the power of her husband Nicolae — Romania's dictator until communism fell — still made sure that the University of Bucharest awarded her a PhD in chemistry. The contents of her many scientific papers were penned by others.

The couple were executed on Christmas Day 1989 for crimes more terrible than poor publication ethics. But their practice of playing fast and loose with academic principles has flourished in notoriously corrupt post-communist Romania.

Over the past 18 months, *Nature* has chronicled an epidemic of plagiarism involving prominent political figures in Europe, reporting, among others, on Germany's former defence minister Karl-Theodor zu Guttenberg, who plagiarized material for his law thesis; Hungary's former president Pál Schmitt, whose thesis on physical education contained plagiarized material; as well as Romania's prime minister Victor Ponta, who continues to dismiss well-founded accusations that he plagiarized sources for his law thesis. On page 264, we highlight widespread plagiarism within Romania's universities and the worrying fact that so many academics there seem not to realize why this is a problem.

Plagiarism seems to be disturbingly prevalent among the European political class and in Romanian academia, but cases continue to pop up everywhere. The Internet makes it easier to detect, but also easier to perpetrate, because anyone can cut and paste more or less anything in a matter of seconds. Perhaps it is this easy access to the words of others that encourages some academics to think that plagiarism is not a serious issue.

They are wrong. Plagiarism is illegal (theft of intellectual property) and immoral, and anyone whose reputation and career rely on publishing their ideas and findings needs to care about it. True, it does not directly affect the scientific literature in the same way as other types of misconduct such as data fabrication. But it has an indirect impact on the academic system because it helps to promote the careers of the fraudulent and the undeserving. And those who climb the academic ladder on the back of dishonest publication records often imbue their students with the same disrespect for scientific method and academic principles.

It matters on a larger scale, too. Most countries accept that to attain economic prosperity they need a robust research base, a concept enshrined in the European Union's Treaty of Lisbon. But a research base contaminated with plagiarism can never function optimally. Romania, a signatory to the treaty, seemed to be on track to a more honest and promising future when it passed its education law, designed to inject competition into its universities and root out widespread scientific misconduct. Yet that law is now being undermined by political interference in the very ethics councils that should be helping to implement it. Against this backdrop, it is easy to see why Romania's excellent scientists — and there are many of them — choose to work mostly outside the country.

A group of researchers is now trying to change things, partly through a website to track and investigate cases of misconduct in Romania. Their stated goal is to "reform and restore confidence" in the country's academic system. Scientists everywhere should back their effort and pass on their message — with appropriate attribution, of course. ■

Safety shambles

Lax management of Fukushima clean-up intensifies concerns over Japan's nuclear future.

Earlier this summer, it emerged that five people working to clean up Japan's devastated Fukushima nuclear power plant had covered their dosimeters with lead. They blocked the radiation detectors — intended to alert them when their exposure was reaching dangerous levels — late last year to allow them to work longer hours.

An announcement made last week by the Tokyo Electric Power Company (TEPCO), which runs the plant, found that the five workers were contracted by a subcontractor of a subcontractor, and had not even been authorized to work at the plant. Other workers were found who were not using the devices at all. In response, TEPCO has barred its direct subcontractor, Tokyo Energy and Systems, from bidding for contracts for three months. The ministry of health and labour is still investigating.

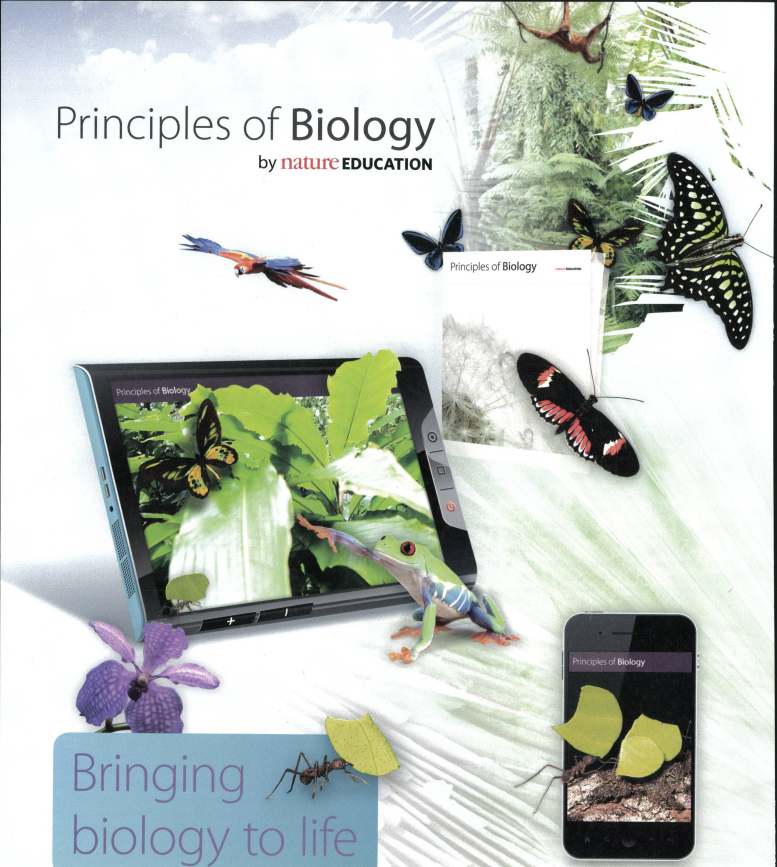
That such egregious flouting of safety protocols would occur despite the media attention on the clean-up efforts is astonishing. And it seems all the more so given the ongoing concern about the health risks of radiation in Japan, and that the nuclear industry's lack of transparency and cavalier attitude towards safety are fuelling a debate over atomic power that is becoming the most politically divisive in the country for decades.

That's on the ground. Decision-making looks no prettier at TEPCO headquarters. Last week, the company made available some 150 hours of video recordings of exchanges between the company and staff at its nuclear plants in the first days of the disaster. The videos highlight some uncomfortable facts that the company had previously denied — that senior officials hesitated to use sea water to cool overheating reactors for fear of damaging them, for example, despite recommendations from a local plant manager to do so.

In the days after 11 March 2011, the world saw that TEPCO's reactors were not designed to withstand disaster. Now it is becoming clear how poorly the company has dealt with the aftermath. ■

Principles of **Biology**

by nature EDUCATION



Bringing biology to life

To learn more about *Principles of Biology*, the first interactive textbook from Nature Education, visit **www.nature.com/principles**

nature publishing group npg

YALE LAW SCHOOL

Why we are poles apart on climate change

The problem isn't the public's reasoning capacity; it's the polluted science-communication environment that drives people apart, says **Dan Kahan**.

Understandably anxious to explain persistent controversy over climate change, the media have discovered a new culprit: the public. By piecing together bits of psychological research, many news reporters, opinion writers and bloggers have concluded that people are simply too irrational to recognize the implications of climate-change science.

This conclusion gets it half right. Studying things from a psychological angle does help to make sense of climate-change scepticism. But the true source of the problem, research suggests, is not that people are irrational. Instead, it is that their reasoning powers have become disabled by a polluted science-communication environment.

Social-science research indicates that people with different cultural values — individualists compared with egalitarians, for example — disagree sharply about how serious a threat climate change is. People with different values draw different inferences from the same evidence. Present them with a PhD scientist who is a member of the US National Academy of Sciences, for example, and they will disagree on whether he really is an 'expert', depending on whether his view matches the dominant view of their cultural group (D. M. Kahan *et al. J. Risk Res.* **14,** 147–174; 2011).

The positions on climate change of both groups track their impressions of recent weather. Yet their impressions of what the recent weather has been are polarized, too, and bear little relationship to reality (K. Goebbert *et al. Weath. Clim. Soc.* **4,** 132–144; 2012). But is this sort of cultural polarization evidence of irrationality? If it is, then how can we explain the fact that members of the lay public who are the most science literate, and the most proficient at technical reasoning, are also the most culturally polarized (D. M. Kahan *et al. Nature Clim. Change* http://dx.doi.org/10.1038/nclimate1547; 2012)?

If anything, social science suggests that citizens are culturally polarized because they are, in fact, too rational — at filtering out information that would drive a wedge between themselves and their peers.

For members of the public, being right or wrong about climate-change science will have no impact. Nothing they do as individual consumers or as individual voters will meaningfully affect the risks posed by climate change. Yet the impact of taking a position that conflicts with their cultural group could be disastrous.

Take a barber in a rural town in South Carolina. Is it a good idea for him to implore his customers to sign a petition urging Congress to take action on climate change? No. If he does, he will find himself out of a job, just as his former congressman, Bob Inglis, did when he himself proposed such action.

Positions on climate change have come to signify the kind of person one is. People whose beliefs are at odds with those of the people with whom they share their basic cultural commitments risk being labelled as weird and obnoxious in the eyes of those on whom they depend for social and financial support.

So, if the cost of having a view of climate change that does not conform with the scientific consensus is zero, and the cost of having a view that is at odds with members of one's cultural community can be high, what is a rational person to do? In that situation, it is perfectly sensible for individuals to be guided by modes of reasoning that connect their beliefs to ones that predominate in their group. Even people of modest scientific literacy will pick up relevant cues. Those who know more and who can reason more analytically will do a still better job, even if their group is wrong on the science.

So whom should we 'blame' for the climate-change crisis? To borrow a phrase, it's the 'science-communication environment, stupid' — not stupid people.

People acquire their scientific knowledge by consulting others who share their values and whom they therefore trust and understand. Usually, this strategy works just fine. We live in a science-communication environment richly stocked with accessible, consequential facts. As a result, groups with different values routinely converge on the best evidence for, say, the value of adding fluoride to water, or the harmlessness of mobile-phone radiation. The trouble starts when this communication environment fills up with toxic partisan meanings — ones that effectively announce that 'if you are one of us, believe this; otherwise, we'll know you are one of them'. In that situation, ordinary individuals' lives will go better if their perceptions of societal risk conform with those of their group.

POSITIONS ON CLIMATE CHANGE HAVE COME TO SIGNIFY THE KIND OF PERSON ONE IS.

Yet when all citizens simultaneously follow this individually rational strategy of belief formation, their collective well-being will certainly suffer. Culturally polarized democracies are less likely to adopt polices that reflect the best available scientific evidence on matters — such as climate change — that profoundly affect their common interests.

Overcoming this dilemma requires collective strategies to protect the quality of the science-communication environment from the pollution of divisive cultural meanings. Psychology — along with anthropology, sociology, political science and economics — will play a part. But to apply the insights that social science has already given us, we will have to be smart enough to avoid reducing what we learn to catchy simplifications. ■

NATURE.COM
Discuss this article online at:
go.nature.com/qhmcdy

Dan Kahan *is the Elizabeth K. Dollard professor of law and professor of psychology at Yale Law School in New Haven, Connecticut.*
e-mail: dmkahan@gmail.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

BEHAVIOURAL GENETICS

Genetics of sexual harassment

A gene important in limb development that is normally suppressed after birth can cause female mice to become sexually aggressive, biting the genitals of males, when it is anomalously expressed in the brain.

Jozsef Zakany and Denis Duboule at the University of Geneva, Switzerland, deleted a tiny section of the mouse genome that includes a cluster of developmental genes called *HoxD*. The deletion resulted in the abnormal expression of one *Hox* gene in small populations of brain cells. This caused females in the sexually receptive phase of their reproductive cycle to hyperactively chase male mice, biting the males' genitals and severely mutilating them.

The finding could open up new approaches for investigating the molecular and cellular basis of female courtship behaviour, say the researchers.

Curr. Biol. **http://dx.doi.org/10.1016/j.cub.2012.06.067 (2012)**

S. BENSON-AMRAM

ANIMAL BEHAVIOUR

Curious hyenas crack puzzles

Spotted hyenas that exhibit a wider range of exploratory behaviours are better at solving problems than those that show a smaller range — just as creativity can aid problem-solving in humans.

Sarah Benson-Amram and Kay Holekamp at Michigan State University in East Lansing devised a steel puzzle box and baited it with a slab of meat (pictured). Over the course of a year, the researchers observed how 62 spotted hyenas (*Crocuta crocuta*) from two clans in Kenya interacted with the box. The team measured the hyenas' reluctance to touch the box — as a proxy for their fear of new things — and the number of tactical behaviours that they used to try to open it, such as biting, pushing or flipping, as well as their persistence. In both adult and juvenile hyenas, animals that showed a broader range of behaviours were more likely to succeed in opening the box.

Proc. R. Soc. B **http://dx.doi.org/10.1098/rspb.2012.1450 (2012)**

NANOTECHNOLOGY

Ramped up resolution

Colour images with an effective resolution of up to 100,000 dots per inch — the maximum possible owing to the properties of visible light — could now be a reality, thanks to researchers in Singapore.

Joel Yang at the Institute of Materials Research and Engineering in Singapore and his colleagues began with a silicon wafer and created tiny posts tens of nanometres in diameter on the surface. The posts were capped with silver disks. Depending on their size and spacing, groups of posts preferentially reflected a single colour back towards the viewer. Using their system, the researchers created colour images (**pictured**), which could be resolved at the diffraction limit of optical microscopes.

The posts could be used in security images and in optical filters or data storage, the team suggests.

Nature Nanotechnol. **http://dx.doi.org/10.1038/nnano.2012.128 (2012)**

For a longer story on this research, see go.nature.com/vwahok

EVOLUTION

Sloth inner-ear diversity

The three-toed sloth shows surprising diversity in the structure of its semicircular canals — part of the inner ear responsible for detecting motion. The sloths' lethargy and lack of agility might have reduced the need for sensitivity to motion, leading to relaxed selection pressure on this apparatus.

Semicircular-canal morphology is usually remarkably stable between members of the same species in placental mammals. Guillaume Billet at the University of Bonn, Germany, and his colleagues used high-resolution computed tomography to compare the semicircular canals of the three-toed sloth (*Bradypus*

variegatus) with those of faster-moving mammals. The structures showed greater variation in shape, size and orientation in sloths than in squirrels, moles or armadillos.
Proc. R. Soc. B **http://dx.doi.org/10.1098/rspb.2012.1212 (2012)**

MATERIALS

Polymers track the Sun

Solar-cell arrays catch more light when they can tilt to follow the Sun's path, but this motion tracking does not have to be driven by power-hungry machinery.

Hongrui Jiang and his colleagues at the University of Wisconsin–Madison mounted a platform of solar cells on top of elastic supporting columns. Each column is coated with films made up of carbon nanotubes embedded in liquid crystal elastomers — rubbery networks composed of ordered polymers. When the nanotubes in the columns nearest the Sun absorb sunlight, they convert the energy to heat, and this changes the orientation of the polymer chains, causing the columns to contract. This contraction tilts the solar-cell platform towards the Sun. As the Sun moves across the sky, different columns contract, changing the platform's tilt to track the Sun's motion.
Adv. Funct. Mater. **http://dx.doi.org/10.1002/adfm.201202038 (2012)**

ZOOLOGY

Beetles walk underwater

A terrestrial leaf beetle can walk underwater, thanks to tiny air bubbles trapped between hair-like structures on its feet.

INT. INST. WATER. SCI.

Gastrophysa viridula (**pictured**) use these 'hairs', or setae, and the secreted fluid that covers them, to stick to leaves when on land. Naoe Hosoda at the National Institute for Materials Science in Tsukuba, Japan, and Stanislav Gorb at the University of Kiel in Germany found that when the animals walk underwater, air bubbles remain trapped between the setae. The bubbles not only provide adhesion but also de-wet the area around the beetles' feet, enabling the setae to function as they do on land.

On the basis of these principles, the researchers were able to develop an artificial polymer that successfully kept small objects attached to vertical surfaces underwater.
Proc. R. Soc. B **http://dx.doi.org/10.1098/rspb.2012.1297 (2012)**
For a video linked to this research, see go.nature.com/ecyztv

REGENERATIVE MEDICINE

Nanofibres foster blood vessels

Self-assembling nanofibres can help to spur the formation of blood vessels in the hearts of rats and pigs during recovery from experimentally induced heart attacks.

Researchers led by Patrick Hsieh of National Cheng Kung University Hospital in Tainan, Taiwan, injected protein fragments that self-assemble into nanofibres, and VEGF — a protein that promotes blood-vessel formation — into the animals' injured heart muscle. Whereas VEGF alone conveyed no noticeable benefits, animals that received both nanofibres and VEGF had more arteries and stronger heart performance four weeks after the simulated heart attacks. The nanofibres, which degrade slowly over time, seem to function as a scaffold that retains and recruits restorative cells.
Sci. Transl. Med. **4, 146ra109 (2012)**

COMMUNITY CHOICE

The most viewed papers in science

COMPUTATIONAL BIOLOGY

'Whole-cell' computer model

HIGHLY READ on www.cell.com 9 Jul–8 Aug

By producing a mathematical model of an entire human bacterial pathogen, researchers have made predictions about its cellular behaviour. The model accounts for the functions of the bacterium's known genes and the interactions between its constituent molecules.

Built by Markus Covert at Stanford University in California and his colleagues, the model captures the life cycle of the bacterium *Mycoplasma genitalium*, which has 525 genes. It incorporates data from more than 1,900 experimental measurements and encapsulates 28 groups of cellular processes, from DNA replication to protein folding. Using their tool, the researchers predicted that, on average, there are more than 30,000 collisions between pairs of DNA-binding proteins on the organism's single chromosome per cell cycle — a number that would be difficult to deduce experimentally.

Whole-cell models could speed up biological research and even pave the way for computer-designed organisms, the researchers say.
Cell **150, 389–401 (2012)**

T. SPADY/OSTRANDER LAB

GENETICS

Heady dog genetics

Variation in a handful of genes explains why some dog breeds, such as pugs, have round, thickset heads, whereas others, such as collies, have long, narrow skulls.

Selective breeding by humans has had a strong influence on the shapes of dogs' heads, which vary greatly across the species. Elaine Ostrander at the National Human Genome Research Institute in Bethesda, Maryland, and her colleagues compared the shapes of 533 skulls belonging to 120 dog breeds (**a sample pictured**) and four subspecies of grey wolf with genetic data from many of the same species. Five genomic regions seem to set breeds with round heads apart from those with elongated heads.

Sequencing one of these regions in 11 dog breeds uncovered a mutation in a developmental gene called *BMP3* in round-headed bulldogs and Pekingese. Further analysis revealed this mutation in nearly all breeds with very short heads, including pugs, Boston terriers and Shih Tzus.
PLoS Genetics **8, e1002849 (2012)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

India's GM caution

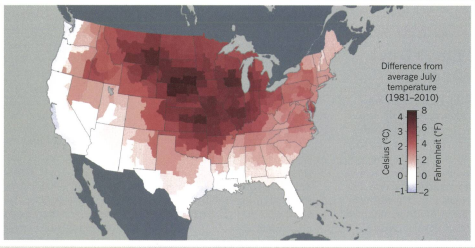
A cross-party panel of politicians in India has urged the government to halt all open-field trials of transgenic crops until it develops a better monitoring and oversight system. In a report released on 9 August, the Parliamentary Standing Committee on Agriculture in India also called for a complete overhaul of the regulatory system for transgenic-crop approvals and recommended that all genetically modified (GM) products be labelled. India has approved only one GM crop (cotton) for cultivation; GM brinjal (aubergine) was approved in 2009, but in 2010 the government banned its cultivation indefinitely, after public opposition. See go.nature.com/r6zcy8 for more.

Development work

The United Nations said on 9 August that it is to set up a global network of research centres, universities and institutes to help identify the best paths to sustainable development. The Sustainable Development Solutions Network will be directed by economist Jeffrey Sachs, who founded the Millennium Villages international development project and heads the Earth Institute at Columbia University in New York City. The move follows on from the UN's Rio+20 conference on sustainable development in June.

Swine-flu alert

The number of reported cases in an outbreak of H3N2v virus — a variant strain of swine flu that can pass from pigs to humans — took a sudden spike last week. The US Centers for Disease Control and Prevention in Atlanta, Georgia, reported 154 cases on 10 August, up from 16 on 3 August; by 13 August the total had reached 176. Most are in Indiana and all are the result of contact with pigs by farmers or through agricultural fairs. Symptoms are mild, and the virus does not seem to transmit between humans. But it is raising eyebrows because it contains a gene from the H1N1 pandemic strain that may increase transmissibility among humans. See go.nature.com/z75z8k for more.



WWW.CLIMATEWATCH.NOAA.GOV/IMAGES

US swelters in hottest month ever

July was the hottest month ever recorded in the United States, the National Oceanic and Atmospheric Administration revealed on 8 August. The average temperature across the continental states was 25.3 °C, one-tenth of a degree above the previous 1936 July record.

Compared with the average for 1981–2010, some parts of the country recorded lower temperatures (blue), whereas parts of the Midwest and the northern Rocky Mountains chalked up readings as high as 4.4 °C above (dark red). See go.nature.com/g94jrm for more.

Russia's space woes

Failures in Russia's space launches have drawn the ire of the country's prime minister. Dmitry Medvedev said on 9 August that the botched launch of two communications satellites earlier that week was costing the country money and prestige, and that he would chair a meeting to look into the failures. Russia lost three navigation satellites in December 2010; a military satellite in February 2011; a telecommunications satellite that August; and the Phobos-Grunt mission to the Martian moon Phobos in November.

EVENTS

Agent Orange

The United States launched an effort on 9 August to clean up dioxin-contaminated soil left over from the Vietnam War at a former American airbase near Da Nang city, Vietnam. The site was one of many hotspots that still have especially high concentrations of dioxins after being sprayed with herbicides such as the infamous Agent Orange (used to defoliate jungle vegetation) from 1962 to 1971. The remediation project, funded by the US government through the US Agency for International Development, is expected to cost US$43 million and take four years.

Iran earthquakes

Two earthquakes of magnitudes 6.3 and 6.4 struck near Ahar in northwestern Iran on 11 August, injuring thousands in rural villages and resulting in a death toll reported at more than 300 people as *Nature* went to press. The region has experienced seven quakes of magnitude 6 or above in the past 40 years, the US Geological Survey says.

BUSINESS

Battery rescue

A123 Systems, a leading US manufacturer of lithium-ion batteries for electric vehicles, may have found a solution to its financial woes. The firm, based in Waltham, Massachusetts,

E. MIN/POLFOTO/PA IMAGES

said on 8 August that it had reached preliminary agreement for an investment of up to US$450 million from Wanxiang Group, China's largest maker of automobile parts, based in Zhejiang. Formed in 2001, A123 is a spin-off from work at the Massachusetts Institute of Technology, Cambridge. In May, it warned investors that it was struggling for cash, and it has now reported net losses of $208 million for the first half of 2012.

Material gains

Oil and gas giant BP will invest US$100 million over ten years in a university-based research centre for advanced materials, it announced on 7 August. With a hub at the University of Manchester, UK, the centre will have smaller UK branches at Imperial College London and the University of Cambridge, and a US branch at the University of Illinois at Urbana-Champaign. It will look at research issues important to BP, such as developing metal alloys that can work in deep-water drilling and membranes for purifying oil, gas, water and biofuels. See go.nature.com/lyzvfe for more.

Drug-discovery split

Irish pharmaceutical company Elan is to split into two, completely separating its drug-discovery work from the main business. The firm, which is based in Dublin, plans to spin off its research work into a small separate company called Neotope Biosciences, which will have its operations base in South San Francisco, California, and will be given at least US$120 million in start-up funding. Elan itself will keep ownership of assets such as its stake in multiple-sclerosis drug Tysabri (natalizumab). The separation, announced on 13 August, should be complete by the end of the year, Elan said.

PEOPLE

Misconduct fall-out

A prominent Danish neuroscientist could lose her PhD and medical-sciences doctorate, after an 8 August report from a panel assembled by University of Copenhagen found evidence that she might have manipulated data in 15 papers. Milena Penkowa (**pictured**), who resigned from the university in December 2010, has denied the accusations in a response appended to the report. The government-run Danish Committees on Scientific Dishonesty will now determine whether Penkowa committed research misconduct, and an academic council at the university will decide whether she can keep

the advanced degrees she earned there. See go.nature.com/adhb1u for more.

RESEARCH

Temperature record

Physicists at the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, say that they have achieved the hottest man-made temperatures ever, by colliding lead ions to create a quark–gluon plasma. Presenting data on 13 August at the Quark Matter 2012 meeting in Washington DC, researchers on the ALICE heavy-ion experiment said that they had created a plasma 38% hotter than the record 4-trillion-degree-Celsius plasma achieved in 2010 by a similar experiment at Brookhaven National Laboratory in New York. The measurement has not yet been converted into a temperature, but could produce something like 5.5 trillion degrees Celsius, says ALICE spokesman Paolo Giubellino. See go.nature.com/kvuq7k for more.

Trust, but verify

An effort to encourage authors of high-profile research papers to get their results replicated by independent labs was launched on 14 August. The Reproducibility Initiative will work through Science Exchange, an online marketplace based in Palo Alto, California, that connects researchers wishing to outsource their experiments with service providers. The initiative will start by accepting 40–50 studies for validation, said Science Exchange. Authors will pay for validation studies themselves, and will be able to publish the results in *PLoS ONE*. See go.nature.com/gc9i9v for more.

COMING UP

17–23 AUGUST
NASA's Mars rover Curiosity will attempt its first drive on the red planet next week — although no firm date had been set as *Nature* went to press.
www.nasa.gov/msl

19–23 AUGUST
The American Chemical Society holds its autumn meeting in Philadelphia, Pennsylvania, focusing on materials for health and medicine.
go.nature.com/tojBlw

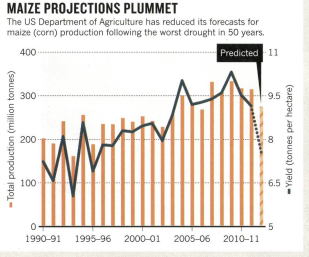
23–25 AUGUST
The brain's connectome — along with other ways of mapping neurons and neuronal circuits — features on the agenda of the 3rd Annual Aspen Brain Forum in Colorado.
http://aspenbrainforum.com

TREND WATCH

The US Department of Agriculture predicted on 10 August that the nation's maize (corn) yields would drop to a 17-year low in 2012–13 (see chart), after a catastrophic drought that has scorched fields and sent food prices soaring. The Food and Agriculture Organization of the United Nations has suggested that the United States suspend biofuel production from maize, and US President Barack Obama unveiled an aid package for stricken farmers on 13 August.

MAIZE PROJECTIONS PLUMMET

The US Department of Agriculture has reduced its forecasts for maize (corn) production following the worst drought in 50 years.



SOURCES: USDA; GO.NATURE.COM/MCHCXN; GO.NATURE.COM/HYZGDF

NATURE.COM
For daily news updates see:
www.nature.com/news

pipetman® L

NEW

Increase Productivity
Without Sacrificing Comfort!

You asked, we delivered. PIPETMAN® L was designed with you in mind pairing legendary PIPETMAN robustness with new features to help you increase productivity and pipette longer without sacrificing comfort.

- *Lockable Volume Ensures No Accidental Volume Changes*
- *Lightweight Body for Comfortable Pipetting All Day*
- *Low Pipetting & Tip Ejection Forces*

www.PipetmanL.com

GILSON®

Tel.++49(0)941 948228 Fax. ++49(0)941 9468119

Davids Biotechnologie

Antibodies made in Germany
Directly from the manufacturer

peptides
rec. proteins

monoclonal

polyclonal

Protein production services

e-mail: info@davids-biotech.de

www.davids-bio.de

production
purification
modification

Davids Biotechnologie, Roentgenstrasse 3, 93055 Regensburg

CUSTOM MONOCLONAL ANTIBODIES
Rat mAb ◦ Mouse mAb ◦ Ascites

CUSTOM PROTEIN EXPRESSION
HEK293 ◦ Baculovirus ◦ *E. coli*

Superior Quality • Unrivaled Expertise
www.promab.com
1.866.339.0871 | info@promab.com

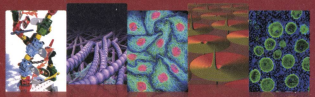
ProMab Biotechnologies, Inc.

nature conferences

Unique events committed to scientific excellence

Sign up for natureconferences e-alerts at
www.nature.com/natureconferences

Receive updates on
- Important deadlines
- Newly added conferences
- Highlights from past conferences



www.nature.com/natureconferences

nature publishing group npg

NEWS IN FOCUS

PLANETARY SCIENCE Curiosity's chief scientist plots Mars exploration **p.262**

MISCONDUCT Romanian academics launch effort to root out plagiarism **p.264**

INFECTIOUS DISEASE Uganda lab is a local hero in Ebola outbreak **p.265**

GENOMICS Silk Road expedition explores the genetics of taste **p.269**

P. PARKS/AFP/GETTY

Protesters in Qidong, Jiangsu province, persuaded officials to scrap plans for a waste pipeline.

ENVIRONMENT

Green protests on the rise in China

Environmental groups use momentum to push for reforms.

BY NATASHA GILBERT

For years, people in China have accepted murky air, tainted waters and scarred landscapes as the unavoidable price of the country's meteoric economic growth. But public dissent over environmental issues has been growing steadily in the communist nation, and now seems to be building the foundations of a fledgling green movement.

In July alone, two separate demonstrations made international news when they turned violent. About 1,000 environmental protesters in the eastern city of Qidong, Jiangsu province, successfully persuaded officials to scrap plans for a pipeline that would have carried toxic waste from a paper-manufacturing plant. The event followed a demonstration earlier that month by several thousand residents in the city of Shifang, Sichuan province, which halted the construction of a copper–molybdenum processing plant. Locals were concerned that the facility would release toxic fumes into the air and generate slag heaps that could leach arsenic into drinking water.

These recent successes come after a slew of ever-larger and more violent green protests over the past few years, as the environmental impact of China's growth becomes harder to ignore. The Chinese public has few legal or regulatory avenues to express concern about environmental damage, forcing people to demonstrate on the streets, says Li Bo, director of Friends of Nature, one of China's first green non-governmental organizations (NGOs), set up in 1993 in Beijing. "These demonstrations are evidence of the public anger and frustration at opaque environmental management and decision-making," says Bo.

Better access to mobile phones and the Internet has enabled protesters to share information and coordinate movements more effectively. The protests also seem to be shifting from rural areas — such as when 500 villagers in eastern Zhejiang province demonstrated in 2011 outside a solar-panel factory that they claimed had discharged waste into local rivers — to cities such as Qidong and Shifang, says Elizabeth Economy, director for Asia Studies at the Council on Foreign Relations, an independent think tank in New York. As the protests have become more visible, widespread news coverage has emboldened others to demonstrate despite police crackdowns, she adds.

Unlike environmental campaigns in the West, however, China's protests rarely have external coordination from green NGOs, says Economy. But some NGOs are now beginning to use the momentum created by the protests to push for regulatory and legal reforms that will improve environmental protections, says Tianjie Ma, who leads a programme on toxic pollution for the East Asia bureau of Greenpeace, the environmental campaigning organization headquartered in Amsterdam.

In what could spark a dramatic shift in green regulation in China, Friends of Nature is suing the Yunnan Liuliang Chemical Industry company for discharging 200,000 tonnes of waste into the Pearl River in the southern province of Yunnan — the first ▶

NATURE.COM
Read more from *Nature China*:
nature.com/nchina

time a grass-roots group has succeeded in bringing a case against a polluter in China. Friends of Nature and other green groups say that the pollution, which contains toxic chromium(VI), may have caused incidents of cancer in local villages. Bo says that the case "is a new initiative to allow pollution victims and entities like NGOs to bring cases to court against polluters". The plaintiffs are demanding that the company pays 10 million renminbi (US$1.6 million) to a compensation fund to clean up the damage.

"There are still lots of issues to be worked out," Bo adds, "such as how to do environmental damage appraisals, and it is still not clear which parties are eligible to be plaintiffs." The court has held two meetings in the past two months to discuss the case, and Bo expects a ruling by the end of the year. "We are hoping this will set a precedent" for other NGOs and members of the public to bring similar cases, he says.

Alex Wang, an environmental lawyer at the University of California, Berkeley, says that the lawsuit signals an opportunity for civil-society groups to play a part in environmental enforcement. But "only time will tell whether this turns into precedent or remains a one-time event. Too many promising initiatives like this one have languished once the initial publicity has died down," he warns.

Green groups are also calling for improvements in the environmental impact assessments (EIAs) that companies must conduct before projects are given permission to go ahead. The public has 14 days to submit concerns and objections to proposed projects in China, but locals rarely have access to detailed plans and EIAs before it is too late, says Ailun Yang, who analyses emerging economies at the World Resources Institute, a think tank in Washington DC.

Yang says that companies should be required to disclose detailed information about the technologies to be used in projects, and to explain how they will dispose of waste. The government should also produce guidance for the public to make it easier for people to participate, she adds.

Keping Ma, a plant ecologist at the Institute of Botany, part of the Chinese Academy of Sciences in Beijing, says that China's central government is developing a "more positive attitude" towards ecological and environmental issues. And local officials did act quickly to cancel the projects in Qidong and Shifang, he points out.

But the strongest pressure for reform comes directly from members of the public, who feel that their lives are threatened by pollution and environmental damage, says Bo. "Throughout several dynasties, China has a history of villages rising up and overthrowing regimes they could not tolerate," he says. "I hope history does not repeat itself, because this has a high cost." ■

PLANETARY SCIENCE

Mars scientists await feast of data

As the Curiosity rover prepares to take its first trip across the surface, the lead scientist shares his hopes for the mission.

BY ERIC HAND

John Grotzinger folds his rangy frame into a plastic leather booth at Conrad's, an unassuming diner in Pasadena, California. It is nearly midnight, and a smattering of customers have drifted in for a late-night snack. In the booth on one side, members of a Neil Young tribute band relax after a gig; on the other, a personal-trainer-cum-actor is winding down from a script reading.

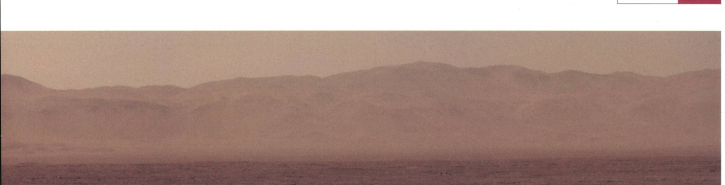
But Grotzinger, who is chief scientist for NASA's Mars Science Laboratory mission, is trying to gear up for the start of his working day at the nearby Jet Propulsion Laboratory, as the rover Curiosity begins its fourth Martian day. Still riding the high of the rover's triumphant landing on 6 August, Grotzinger says that he is "overjoyed" with the landing site, a level plain between the rim of Gale Crater and Aeolis Mons, the imposing mountain at the crater's centre. But he is sobered by the demands of managing such a complicated machine and the 400-strong science team behind it. "I feel the burden of two-and-a-half billion dollars," says Grotzinger, a geologist at the California Institute of Technology in Pasadena. "I feel the burden of the future of Mars exploration."

Grotzinger's job is to get the maximum scientific pay-off from that investment during the two years or more that the rover is expected to remain active. His day will begin at 12:42 a.m. with a science discussion with the principal investigators for the rover's ten instruments. But first Grotzinger needs breakfast and some coffee. One cup will do: the excitement of new data, which are expected in a downlink in two hours' time, will be enough to keep him alert.

Grotzinger was a regular at Conrad's in 2004, before and after his working days on the rover Opportunity, which landed that year along with Spirit, its twin, comprising the Mars Exploration Rover mission. Because the rovers were positioned on opposite sides of Mars, one team would be having breakfast while the other would be eating dinner. "The waitresses were always confused," he recalls. This time there is only one rover, but still no standard working day. Adapting to 'Mars time' requires starting each Earth day 40 minutes later than the last to match Martian daylight, inducing a state of perpetual jet lag.

Grotzinger is already marshalling different opinions on where the rover should go for its inaugural drive, expected to take place in the next week. The first images from the rover's navigation cameras suggested one possibility when they revealed

NATURE.COM
For more on Curiosity, see:
nature.com/curiosity



The first high-resolution colour panorama from Curiosity's mast camera reveals the distant rim of Gale Crater.

NASA/JPL-CALTECH/MSSS

whitish linear features in a shallow trench scoured out by the blast of the landing system's retrorockets. The lines could be fractures filled with water-altered minerals — key targets for exploration in the mission's goal of assessing whether ancient Mars could have supported life. "We want to find out if we've got a bird in the bush right there," Grotzinger says.

Several hundred metres away is a second option for exploration: a five-metre-high slope, or scarp, that separates the dusty desert pavement on which the rover landed from the more hummocky terrain that lies beyond. The scarp marks the edge of materials deposited at the foot of an alluvial fan, a sedimentary feature left behind from floods that once spilled down from the rim of Gale Crater. The materials seem to be cemented or consolidated in some way, and layers exposed on the scarp might reveal the comings and goings of an ancient lake.

Away in the distance is the rover's ultimate destination: the 5.5-kilometre-high Aeolis Mons, dubbed Mount Sharp by the science team, which casts a long shadow over the rover each morning. Data from orbiting spacecraft suggest that deposits on the mountain's lower shoulders were formed in water — and for now at least, these deposits represent the team's best chance of working out whether microbes could have lived on Mars billions of years ago.

Managing the tension between short-term and long-term goals will be a constant challenge for Grotzinger, says Raymond Arvidson, deputy principal investigator for the Spirit and Opportunity rovers and a planetary scientist at Washington University in St Louis, Missouri. "Do we stop and smell the roses or do we button up and continue to the next outcrop?"

Colleagues say that, in his geological field studies on Earth, Grotzinger balances both tendencies, displaying an exuberant fitness that carries him across wide terrain even as he keeps an eye on the minutiae of mineralogy and texture. He scrambles like a mountain goat, says Ralph Milliken, a former postdoctoral student of Grotzinger's and now a planetary scientist at Brown University in Providence, Rhode Island. Yet Andrew Knoll, a palaeontologist at Harvard University in Cambridge, Massachusetts, recalls a field trip with Grotzinger to Siberia in the early 1990s, during which the geologist discovered small pieces of volcanic breccia rock underfoot that had gone unnoticed by Knoll, no slouch of a field geologist himself. By dating those rocks, the researchers were able to pin down the timing of the Cambrian boundary better than ever before, and show that the Cambrian explosion of animal life happened incredibly quickly (S. A. Bowring *et al. Science* **261,** 1293–1298; 1993).

Having an eye for detail and a gift for stratigraphic field work will certainly help Grotzinger to decode Mount Sharp and the millions of years of history that its layers represent. But he faces sociological challenges as well as scientific ones. Unlike Steven Squyres, the charismatic planetary geologist from Cornell University in Ithaca, New York, who was principal investigator for Spirit and Opportunity, Grotzinger does not have final authority over plans for the rover. Instead, as project scientist, he heads an executive committee that is made up of a representative from NASA headquarters and the principal investigators for the ten instruments. "He appreciates that he's the head of a committee, and not the Pope," says Knoll.

Arvidson says that Grotzinger is already aware of the job's limitations. "John's working the crowd," he says. "He really understands that in order to get buy-in, you have to establish trust."

One exercise that Grotzinger is proud of is a crowd-sourced geological map of Curiosity's landing ellipse that was put together before touchdown. Recognizing that little was known about the fine-grained detail of the location, he divided the ellipse into 151 quadrangles, measuring roughly 1.6 kilometres on each side, and asked for volunteers from the science team to turn orbiter data into geological maps of each sector. It was partly a team-building effort in the lead up to the landing — how better to bring opinionated geologists into the fold than by working on a patchwork quilt? But he also wanted to show the wider team that the landing terrain was full of intriguing detail that should not be overlooked in a fast trek to Mount Sharp.

"There is going to be something for everyone," says Grotzinger, slurping his coffee and smiling. "I quoted Bob Dylan when I was talking to these guys … 'A lot of people don't have much food on their table. But they got a lot of forks 'n' knives. And they gotta cut somethin.' If you have an empty plate, people are going to go for each other. But if there's a wealth of riches, we can all share." ■

P. E. ALERS/NASA

Geologist John Grotzinger is plotting a course for Curiosity.

S. JANOS/REX FEATURES

The university at Târgu-Mureș is at the centre of plagiarism accusations.

MISCONDUCT

Romanian scientists fight plagiarism

Researchers set up independent review panel after misconduct scandals hit government.

BY ALISON ABBOTT

Plagiarism scandals have shaken Romania this year. Former research minister Ioan Mang[1] and current Prime Minister Victor Ponta[2] have both been implicated, and leaders at a large medical university in Târgu-Mureș face similar accusations — although the government seems to have ignored those charges. Such scandals have convinced some Romanian scientists that they need to fight back against a culture of plagiarism that they see as ingrained in the university system.

This week, researchers are launching an online service called Integru, which will investigate and expose cases of plagiarism and other academic misconduct in Romania. Each case will be accompanied online by commentaries from international — and independent — reviewers selected for their expertise in the relevant field.

Integru's stated goal is to "help reform and restore confidence in the Romanian research and education system". Its editorial board, which comprises Romanian scientists working inside and outside the country, will remain anonymous to avoid personal attacks from those accused of misconduct.

Some Romanian academics facing allegations of plagiarism "have tried to dodge the bullet by questioning the motivation of the accusers", says one Integru board member. For example, both Mang and Ponta have belittled the charges against them by arguing that the whistle-blowers were politically motivated. Integru should be able to thwart that strategy, says the board member, "because suspicious papers will be posted alongside documentation of plagiarized and copy–pasted sections, and commented on by credible named experts from credible academic institutes around the world — the misconduct will be hard to refute".

Scientists concerned about plagiarism had once pinned their hopes on Romania's National Ethics Council, which last year was given a strengthened mandate to investigate wrongdoing under a law aiming to raise standards in universities. The council published decisions on 15 cases, and reached conclusions on another three. But many lost faith in its fairness when Ponta's government dismissed all the council's members on 8 June, just as they were poised to vote on the plagiarism charge against Mang (who had already resigned his ministerial post). The council was reconstituted with new members appointed by the government, and its first act was to clear Ponta of his plagiarism charges in July. However, an ethics commission established by the University of Bucharest, which had awarded Ponta his PhD, contradicted that verdict, saying that the prime minister was guilty as charged[3].

On the day it was dissolved, the former research-ethics council submitted to the science ministry its report on charges of plagiarism against three leading figures at the University of Medicine and Pharmacy of Târgu-Mureș: rector Leonard Azamfirei; former rector Constantin Copotoiu, who is now president of the university senate; and senate member Klara Brânzaniuc, a former vice-rector for research. The ministry has not published the report and did not respond to *Nature*'s request for information about it.

Marius Echim, a physicist at the Belgian Institute for Space Aeronomy in Brussels and vice-president of the dissolved council, chaired the investigation. He says that it "identified clear plagiarism in all three cases and recommended that two of the three professors involved be dismissed from the university". No action has been taken, and the apparent suppression of the case is disturbing, he says. While Echim was on the council, he adds, it saw cases of extensive plagiarism across many different disciplines in Romania.

The Târgu-Mureș cases refer to papers in the university's English-language journal *Acta Medica Marisiensis*, which contain sections identical to other researchers' works. Stefan-Virgil Hobai, a biochemist who last year retired from the university, e-mailed university faculty members in December 2011 with evidence of

MORE ONLINE

TOP STORY

L. BRITT/ALAMY

A tree's leaves are genetically different from its roots **go.nature.com/gyq9rf**

MORE NEWS

- Shark-tooth weapons reveal lost biodiversity **go.nature.com/acxgoc**
- Quantum teleportation achieved over record distances **go.nature.com/9h2ie4**
- Seizure detector treats epilepsy in rats **go.nature.com/vbyquf**

FROM THE BLOG

A. BARNA/USF PWS

Prisoners pitch in to help save endangered butterfly **go.nature.com/4mmckf**

the plagiarism, which *Nature* has seen. Since then, he has identified a total of 22 cases of apparent plagiarism in *Acta Medica Marisiensis* and 14 cases in another of the university's journals, the English-language *Romanian Review of Laboratory Medicine*. That represents 7% of the papers he examined in each journal.

Peter Hantz, a Romanian physicist of Hungarian ethnicity based at the Friedrich Miescher Institute in Basel, Switzerland, submitted the Copotoiu, Azamfirei and Brânzaniuc cases to the National Ethics Council in January. Azamfirei told *Nature* that he thought Hantz, a prominent campaigner for Hungarian-language teaching in Romania, may have been trying to discredit the ethnic Romanian university leaders. Târgu-Mureș has a large Hungarian population and a history of ethnic tensions.

Sanda-Maria Copotoiu, editor-in-chief of *Acta Medica Marisiensis* (and the wife of Constantin Copotoiu), responded to the allegations by publishing an 'expression of concern' in the latest issue of the journal, referring anonymously to Hobai as a "character who styled himself as a prominent researcher and voluntary justice maker", whose "vicious" allegations are "intoxicating the net". She told *Nature* that all three accused researchers had asked her to remove their names from the papers long before the plagiarism charges reached the ethics council, because they became aware of the papers' existence only after publication. Their names have not yet been withdrawn, but Constantin Copotoiu's paper has been removed from the journal's website.

Constantin Copotoiu and Brânzaniuc did not respond to *Nature*'s request for comment. Azamfirei told *Nature* that there is no plagiarism in the article with his name on it, and that the National Ethics Council was not the appropriate body to review his case because he was not rector at the time the article was published.

Integru asked three reviewers to take a closer look at the Copotoiu case, and all concluded that the article did contain sections of copied text. One of the reviewers, Mihai Ionac, a vascular surgeon at the Victor Babes University of Medicine and Pharmacy in Timisoara, Romania, says that too many researchers in Romania do not take plagiarism seriously, and may not understand why it is classed as misconduct. The Integru initiative is "extraordinary, and comes at the right time, with all the scandals and turmoil now going on", he says. "Just a few years ago, no-one was even talking about plagiarism — it was covered up, like a cat covers up its dirt." ■ SEE EDITORIAL P.253

1. Abbott, A. *Nature* **485,** 289 (2012).
2. Schiermeier, Q. *Nature* **486,** 305 (2012).
3. Schiermeier, Q. *Nature* http://dx.doi.org/10.1038/nature.2012.11047 (2012).

REUTERS/C. EDWALU

Workers at a biosafety-level-3 lab in Entebbe diagnosed recent cases of Ebola.

INFECTIOUS DISEASES

Ebola outbreak tests local surveillance

Ugandan laboratory helps to minimize number of cases.

BY EWEN CALLAWAY

Film-goers know the script: a frightening infectious disease emerges somewhere in the developing world. Masked healthcare workers arrive, blood samples are rushed to an airfield and the grim diagnosis is finally made in a state-of-the art laboratory on a different continent. But the response to the Ebola infections that have killed at least 16 people in Uganda since early July departed from the script — a laboratory in Entebbe, near Uganda's capital Kampala, has passed its first major test after taking a leading role in confirming and tracking the infection.

The lab, run by the US Centers for Disease Control and Prevention (CDC) and based at the Uganda Virus Research Institute (UVRI), tests tissue samples from patients with symptoms of the viral haemorrhagic fever. Its proximity to the outbreak (see 'Local hero') has helped to speed up the response, and the outbreak now seems to be under control. The episode highlights the advantages of having sophisticated surveillance labs in parts of the world where infections are likely to emerge, says Eddie Holmes, a geneticist who studies emerging infections at Pennsylvania State University in University Park. "The CDC making a commitment to this lab is a tremendous step forward."

The laboratory was set up after a 2007 outbreak of a newly discovered strain of Ebola called Bundibugyo, which killed 42 people. The area hit by the outbreak had no electricity and lacked the facilities to house a mobile containment lab, so scientists studying the disease set up shop at the UVRI. After the Bundibugyo outbreak subsided in 2008, the CDC invested in a permanent laboratory there that could diagnose haemorrhagic fevers such as Ebola and Marburg. Its biosafety-level-3 laboratory is safe enough to diagnose these viruses from tissue samples, but not to grow virus-infected cells for more sophisticated analyses, such as genome sequencing to study virus evolution. For that, samples are still sent to the CDC's headquarters in Atlanta, Georgia.

When the Entebbe lab isn't in outbreak mode — samples still arrive daily from the country's midwestern Kibale district — it can provide negative diagnoses for diseases with symptoms similar to those of viral haemorrhagic fever, such as bloody diarrhoea caused by shigellosis.

Uganda has had three Ebola outbreaks since 2000, so the lab should also help to catch future outbreaks early. "The earlier you catch an epidemic happening, the better you can stop its spread," explains Joseph Fair, vice-president of Metabiota, a company based in San ▶

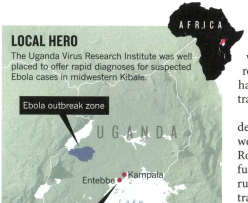
Francisco, California, that conducts research on emerging diseases for the US government.

In May 2011, for example, a 12-year-old girl in Uganda was admitted to hospital with a fever and bleeding from nearly every orifice. She died within hours, and doctors suspected she had been infected with a virus such as Ebola or Marburg. In the past, a conclusive diagnosis would have meant chartering a plane to airlift samples to high-containment labs across Africa or in the United States, which can take weeks. Instead, a blood sample from the girl took just three days to arrive at the Entebbe lab, which confirmed Ebola. Health-care workers reached the girl's village a day later and helped to nip a would-be outbreak in the bud by tracking anyone who might have been exposed (T. Shoemaker *et al. Emerg. Infect. Dis.* http://doi.org/h53; 2012).

Without firm diagnoses, workers tracking outbreaks have to keep tabs on thousands of suspected or potential cases. But having the CDC lab so close to the outbreak means that "we can go back and forth to the site of the epidemic every day", says Pierre Rollin, a virologist at the CDC's Special Pathogens Branch in Atlanta, who is currently in Uganda. "Any samples collected this morning in Kibale will arrive at the lab and be tested tomorrow." In the latest outbreak, rapid diagnosis has meant that health-care workers had to track mere dozens of suspected cases.

"Other countries are asking us if we can develop the same kind of lab for them, but we don't have the resources to do it," says Rollin. Instead, the CDC and other Western funders support programmes to build and run laboratories in the developing world and train their staff, so that countries can monitor and respond to outbreaks with their own resources. Gabon, for example, operates a maximum biosafety-level-4 laboratory.

"You want to build capacity in the country," says Stephen Morse, an epidemiologist at Columbia University in New York, "and not wait until they ask for help to send people parachuting in." ■



CHEMISTRY

Amino acid provides shortcut to drugs

Organocatalyst halves synthesis of prostaglandin family.

BY KATHARINE SANDERSON

An iconic but convoluted pharmaceutical synthesis received a major overhaul this week. The approach, described online in *Nature*, gives researchers easier access to a family of molecules that includes a blockbuster drug.

The efficient new route to prostaglandins, a class of hormone-like chemicals that influence processes including blood circulation and inflammation, offers a way to churn out many different versions of the molecules. It also marks a victory for one of the hottest fields in organic chemistry: organocatalysis, which speeds up chemical reactions with the help of carbon-based molecules, potentially avoiding some of the drawbacks of conventional metal catalysts.

Prostaglandin-like drugs are tempting targets for medicinal chemists; one example, Pfizer's glaucoma drug latanoprost, achieved annual sales of around US$1.75 billion before its patent expired in 2011. But it takes an arduous 20 chemical steps to make latanoprost using the original strategy developed in 1969 by Elias Corey at Harvard University in Cambridge, Massachusetts. The synthesis helped Corey to win the Nobel Prize in Chemistry in 1990.

Synthetic chemist Varinder Aggarwal at the University of Bristol, UK, realized that he could radically simplify the process by building the heart of a prostaglandin molecule in a different way. He applied a workhorse of organic chemistry, the aldol reaction, to build key carbon–carbon bonds in a prostaglandin's five-atom ring, but used the amino acid (*S*)-proline as an organocatalyst to make sure that the end product had precisely the right shape (G. Coulthard *et al. Nature* http://dx.doi.org/10.1038/nature11411; 2012). Corey's approach takes nine steps to make a different intermediate with the same crucial ring.

The molecule produced in Aggarwal's reaction serves as a foundation stone for the whole family of prostaglandins, and his method can produce grams of this building block at a time — a viable scale for drug-development work. Aggarwal and his colleagues used it to make the most complicated prostaglandin, $PGF_{2\alpha}$, which is very similar to latanoprost, in just seven steps (see 'On target'). It's "by far the single best approach to the prostaglandins", says Ben List, an expert in organocatalysis at the Max Planck Institute for Coal Research in Mülheim an der Ruhr, Germany. As well as being shorter than Corey's version, it is very specific at producing only left-handed or right-handed versions of the molecules — a crucial factor in making an effective drug.

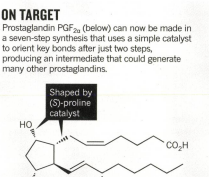
Whether the synthesis helps to deliver new drugs remains to be seen, says biomedical researcher Colin Funk at Queen's University in Kingston, Ontario, Canada. "Prostaglandin-like drugs are still relevant, but the big thing has been to try and limit their side effects," he says. That demands a better biological understanding, rather than a new synthetic route, he notes.

The method might prove useful in full-scale drug production, not only because of its simplicity but because it avoids expensive and toxic metal catalysts. "From an industrial perspective Aggarwal's process is promising, and a remarkable lead," says process chemist Rob Singer, who works at Pfizer Worldwide Research and Development in Groton, Connecticut. The synthesis would still need further development before it could be used at a production scale, he adds.

List says that most pharmaceutical companies are studying organocatalysts, but so far have tended to use them on only a small scale. As the chemical structures of drugs get more complicated, he hopes that organocatalysts will gain more prominence, and Aggarwal's work should help to speed the change, he says. "It reflects the way that organocatalysis fundamentally changes the way we do catalysis," he says. ■

ON TARGET

Prostaglandin $PGF_{2\alpha}$ (below) can now be made in a seven-step synthesis that uses a simple catalyst to orient key bonds after just two steps, producing an intermediate that could generate many other prostaglandins.



AIDS

Generic HIV drugs will widen US treatment net

Upcoming patent expiries stand to make medicines cheaper — although less convenient.

BY AMY MAXMEN

HIV remains a devastating scourge worldwide — and an expensive one. In the United States, where HIV care costs the most, the annual price of treatment averages almost US$20,000 per patient, and the federal government foots most of the bill. This year alone, the United States will spend nearly $15 billion on HIV care and medication, yet funding shortfalls mean that more than 2,000 patients remain on a waiting list for antiretrovirals, and many others are not receiving treatment for co-infections such as hepatitis C.

In the next few years, that economic burden looks set to ease as US patents on the 'superstar' HIV drugs begin to expire (see 'Time's up'). Generic versions of brand-name antiretrovirals, whose use is permitted in the developing world, could reduce the cost per patient and allow the US budget to stretch further. "AIDS, as we currently know it, may become much more affordable," says John Bartlett, an infectious-disease researcher at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland.

Earlier this year, the HIV/AIDS working group at the Department of Health and Human Services asked Rochelle Walensky — an infectious-disease physician at Massachusetts General Hospital in Boston who is known for her cost-effectiveness analyses — for an opinion on the impact of a switch from Atripla, a widely used 3-in-1 HIV pill, to a combination of generic and brand-name drugs might be. Atripla is the first of four first-line drug combinations recommended by US guidelines that will be affected by patent expiries. A key component of the pill will lose protection next year.

At last month's International AIDS Conference in Washington DC, Walensky announced that the switch would save the United States a projected $920 million in the first year alone. She estimates that Atripla — a co-formulated pill from Gilead Sciences in Foster City, California, that consists of brand-name emtricitabine, efavirenz and tenofovir — costs about 40% more than an alternative three-pill regimen. This would include: generic lamivudine, which can be used as a substitute for emtricitabine; a generic version of efavirenz that is expected to hit the market once the brand goes off-patent in 2013; and brand-name tenofovir.

In her presentation, Walensky estimated that the replacement could be slightly less effective than Atripla, not least because patients are more likely to skip doses when they have to swallow several pills. But the difference is small, according to Walensky's model: staying with Atripla rather than switching to the three-pill regimen would add 4.4 months to a patient's life, but $42,500 to their bill.

TIME'S UP
Patents on several key HIV drugs are set to expire over the next five years

Drug	Type	Manufacturer	Expiry
3TC (lamivudine)	Nucleoside reverse transcriptase inhibitor	GlaxoSmithKline	2010
Abacavir	Nucleoside reverse transcriptase inhibitor	GlaxoSmithKline	2012
Efavirenz*	Non-nucleoside reverse transcriptase inhibitor	Bristol-Myers Squibb	2013
Delavirdine	Non-nucleoside reverse transcriptase inhibitor	Pfizer	2013
Darunavir	Protease inhibitor	Tibotec	2015
FTC (emtricitabine)*	Nucleoside reverse transcriptase inhibitor	Gilead	2015
Tipranavir	Protease inhibitor	Boehringer Ingelheim	2015
Ritonavir	Protease inhibitor	Abbott	2016
Tenofovir*	Nucleotide reverse transcriptase inhibitor	Gilead	2017

*Component of Atripla

Even in the United States, where discussions of cost-effectiveness in health care are politically charged, this cost-benefit ratio may be too high for any health-care innovation. "I'm a doctor and I would never want to put a patient on a lesser drug — but we do have to think about trade-offs," Walensky says. In this case, she says, the trade-off could be between a slightly less-effective treatment for the individual but greater care for people with HIV overall.

Patient groups are not immune to this reasoning. "Although there's a convenience factor to having one pill once a day, if the trade-off is many people having no access to medication, then it's better to split up the pill or use the threat of splitting it to get Gilead to reduce their prices," says Michael Weinstein, president of the AIDS Healthcare Foundation, an advocacy organization based in Los Angeles, California. Gilead would not comment directly on pricing, but a spokeswoman told *Nature*, "We believe there is a lot of data supporting the value of single-tablet regimens".

Analogous questions will arise as several other first-line medicines lose patent protection over the next five years. If generic versions are embraced in the United States, drug companies will feel a squeeze that they say could inhibit the development of new HIV drugs. But some say a widespread switch is unlikely to happen, noting that newer patented drugs are on the way, including Gilead's Quad, a combination therapy slated for approval by the US Food and Drug Administration on 27 August. This 4-in-1 pill contains a novel integrase inhibitor, and outperformed Atripla in clinical trials. If the bulk of patients switch directly from Atripla to Quad, Gilead will retain its hold on the US market.

Bartlett argues, however, that the semi-generic substitute may become popular among the thousands of people who currently control HIV with Atripla, with Quad reserved for those in need of an alternative. With so many drugs going off-patent, he predicts, "In ten years this will be a disease treated for $200 per year, or less". ■

CORRECTIONS

The graph in the News Feature 'History as science' (*Nature* **488,** 24–26; 2012) accidentally swapped the labels for 'Racially motivated' and 'Political'. The corrected graph is online at go.nature.com/p1rqsj.

The Feature 'Small, furry ... and smart' (*Nature* **461,** 862–864; 2009) said that Solomon Shereshevsky memorized Dante's *Divine Comedy* from a single read. In fact, he memorized several stanzas of the poem in Italian (a language unfamiliar to him).

Here, the curious will discover.

The curious start with ProQuest Dialog.™ Imagine having information from the world's most authoritative publishers, the workflow tools to search, organize, share and manage it in a collaborative way with speed and precision, and the added value of excellent customer support—it's all yours with the new ProQuest Dialog service. Because, when you have the right information, there are no limits on what you can do.

Dialog®

Discover. Validate. Market. Innovate.

Learn more | **www.dialog.com/proquestdialog**

DialogLLC

TASTE FOR ADVENTURE

Scientists have travelled the ancient Silk Road to understand how genes shape people's love for foods.

BY ALISON ABBOTT

It wasn't quite the lowest point of the expedition. But Paolo Gasparini was decidedly uncomfortable at a banquet in the small town of Alga, Kazakhstan, when he was given the honour of dividing a boiled sheep's head among the gathering. He could sense the silent alarm of his young colleagues, fearing that soon they would be politely chewing on an eye or an ear. Fighting nausea, Gasparini, a medical geneticist at the University of Trieste, Italy, sympathetically sliced each of them just a taste of skin, passing the traditionally valued sense organs to the village elders.

The sheep's-head ritual did, however, neatly reflect the mission of the trip. The Italian scientists were gathering DNA samples from members of the many different ethnic groups that live along the Silk Road — a 2,000-year-old trading route between China and Europe — to map the genetic roots of food preference and of the senses that contribute to it. The team set out from Trieste in July 2010 and collected its final samples this June in northern Kyrgyzstan. It had amassed about 1,100 DNA samples and thousands of sensory-test results from a dozen or so distinct populations, ranging from Caucasian to Turkish and Mongolian.

FULL FLAVOUR

Over the past two decades, researchers have identified receptors on the tongue for sweet, umami ('savoury'), bitter, sour and salty. They have realized that the sensation of flavour is created when signals from these receptors are integrated in the brain with food-related signals from the hundreds of odorant receptors in the nose and nasal passages. But understanding why people prefer certain foods is still a major scientific puzzle.

Researchers have studied genes and food preferences in populations from Georgia to Kyrgyzstan (above). LISA ZILLIO/SISSA MEDIALAB

Subtle variations in the genes that encode the taste and smell receptors are likely to have a big influence. But many genes involved in touch, hearing and sight — those that respond to the creamy feel of fatty food, the sound of crunching nuts, the sight of a crimson tomato — also have a role. Travelling the Silk Road gave Gasparini and his team a unique opportunity to identify genetic influences on food preference by studying isolated populations, in which such variants are relatively easy to find. "The expedition approaches an almost unexplored

dimension of what makes people different," says Yurii Aulchenko, a statistical geneticist from the Russian Academy of Sciences Institute of Cytology and Genetics in Novosibirsk.

Researchers in evolutionary and population genetics are watching with interest, as is the food industry. Humans evolved in environments where food was scarce and poisonous plants a constant threat. As a result, calorie-dense fatty foods can be hard to resist, whereas the bitter tastes that are common in toxic plants can make some nutritious foods, such as broccoli, unpalatable. In the modern, food-rich world, that genetic legacy helps to push people towards diets that contribute to obesity and ill-health. Understanding the roots of food preferences might one day help to change people's behaviour. "Eventually, industry might be able to develop foods that are enjoyable without being bad for you," says Jim Kaput, head of clinical translational research at the Nestlé Institute of Health Sciences in Lausanne, Switzerland, which collaborates with the Nestlé Research Center in Lausanne on studies linking genomics and food preferences. "But right now we understand almost nothing — we need a lot more basic research," says Kaput.

GENETIC GASTRONOMY

The Silk Road expedition — officially called Marco Polo, after the Venetian explorer who travelled the route in the thirteenth century — was born five years ago, when Gasparini met Enrico Balli, director of the Medialab of the International School for Advanced Studies in Trieste. The two hit it off immediately. Balli, a straight-talking physicist turned science-communicator, was itching to film exotic locations. Gasparini, tall and genial, had spent most of his scientific life studying the genetics of isolated communities in dead-end valleys of the Italian Alps. He had become interested in the genetics of food preferences, was consulting with food companies, and had been dreaming vaguely of accessing remote, isolated populations in more grandiose mountain chains in Asia or South America, where scientists have rarely ventured.

Such populations were often founded by just a few people who went on to marry within the group, so their descendants tend to be genetically homogeneous. This makes it easier to work out if a gene variant is associated with a particular characteristic, such as a taste preference, than it is in a large, genetically diverse group. Comparing different isolated populations can be even more valuable. If a variant is associated with the same trait in genetically different groups, that may be a sign that it is important across almost all populations.

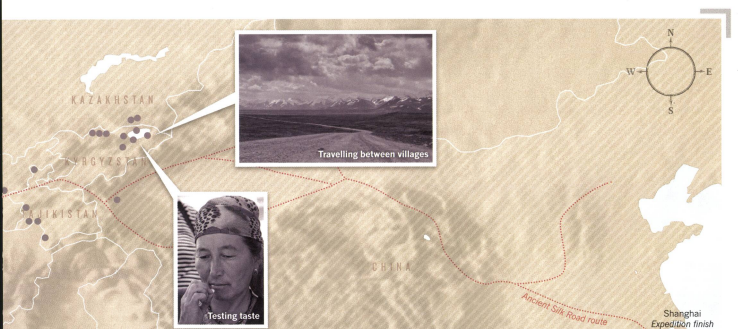
The Silk Road offers a potential paradise for such genetic exploration. The route traverses massive mountain ranges such as the Pamir and the Tian Shan in central Asia and passes through pockets of the nomadic tribes who originally populated the region, as well as ethnically diverse groups descended from traders who settled en route, often near the roadside inns called caravanserais. These populations did not tend to share their genes, but they did share recipes. Cuisines are remarkably similar along much of the Silk Road — variations on tandoor breads, noodles with vegetable or mutton sauces, and dried or fresh fruit. This means that differences in food preferences between groups are likely to be down to variations in genes rather than in dietary cultures, making them even more appealing to the geneticists.

But before Gasparini and Balli could study these groups, they had to reach them. They originally conceived the expedition as a do-it-yourself venture, in which they would simply drive themselves from country to country. Fortunately, they consulted Lilia Smelkova — co-founder of Terra Madre, a network of farmers in developing countries established by the non-profit campaign group Slow Food, which is based near Turin, Italy. Smelkova is an expert in the former Soviet republics that Gasparini and Balli planned to visit. "I thought they were quite crazy," she says, knowing the civil-war-torn, multilingual, sometimes despotic territories that the scientists planned to traverse and the closed communities that they wanted to study. "But I loved their ideas, so I agreed to help — by joining them." Smelkova used her contacts to help the researchers to win the confidence, and the cooperation, of villages on the route.

TASTE TEST

The team, consisting of ten or so people, split the expedition across three summers (see 'Silk-Road science'). The first year, they set off for Georgia armed with just a few clothes and some compact equipment, and fell into a gruelling schedule. Each day, they drove — often for many hours — to a community that was to be sampled and tested. A local Terra

SALIVA: PREFERENCES; TASTE; TRAVELLING: USA-ZULO-SISSA-MEDIALAB; SMELL; HEARING: SISSA-MEDIALAB



Madre helper explained the tests, and volunteers were asked to donate DNA by spitting into plastic tubes. Then the scientists tested volunteers' abilities to hear different sound frequencies, to taste salty or bitter compounds soaked into scraps of filter paper, to distinguish different shades of colour and to identify 12 different smells impregnated in marker-pen-like sticks. The team also showed the volunteers pictures of 80 different food types and asked them to indicate how much they liked or disliked each one. At night, the scientists would feed the results into their computers before collapsing into bed.

The expedition's success depended entirely on getting sample tubes containing saliva back to Italy, but this proved nerve-rackingly difficult. When they arrived home after the first leg in September 2010, the researchers found that Uzbek customs had stopped a precious shipment of 350 tubes gathered in Azerbaijan, Turkmenistan and Uzbekistan. Despite exhaustion, Balli says, "I flew back to Tashkent, and brought them straight back in a suitcase."

Systematic analyses of the Silk Road data will keep the scientists busy for years. The team is carrying out a detailed analysis of small variants known as single-nucleotide polymorphisms (SNPs) in all the DNA samples, and will fully sequence up to 50 samples.

The scientists have already identified eight variants in known genes, including one for an ion channel involved in sensing spicy-hotness, which are associated with a taste for particular foods. And they have found that variants of the gene for the TAS1R2 protein, part of a sweetness receptor, are associated with a strong liking for vodka and white wine (N. Pirastu *et al. J. Food Sci.*; in the press). The results fit with reports linking preferences for alcohol and sweetness in mice (S. M. Brasser *et al. Physiol. Genomics* **41**, 232–243; 2010), and hint that both substances are perceived by the same receptor.

NATURE.COM
For a film clip about the expedition, see:
go.nature.com/ql3eln

There may be bigger scientific stories hiding in the data. Gasparini says that the team is seeing an emerging association in Tajikistani populations between an olfactory receptor gene and both sensitivity to bitter tastes and a tendency to mistake smells. If the finding holds up, it will be the first demonstrated genetic link between smell and taste perception, and it could help to explain how signals from different senses combine to sculpt individual food preferences. The finding may also offer clues to how evolution might have shaped the senses. "Who knows?" Gasparini muses. "Might evolution have compensated those who can't distinguish smells efficiently by selecting genes that make them more sensitive to bitter tastes, so they have a better chance of recognizing poisons?"

DIET AND DIVERSITY

The genetic data will serve another purpose by filling a gap in the international 1000 Genomes Project, which is cataloguing genetic variants in different populations across the globe, but until now had no representatives from central Asia. Population geneticists curious about how different peoples evolved will find their own stories in the Silk Road DNA, says Aulchenko. "We know a lot about the big migrations — out of Africa, for example — but we don't know how trade shaped genetic variation."

The data were hard won. At the trip's lowest point, on the drive through Tajikistan from Qal'aikhum to Khorugh in the Pamir Mountains, a road closure forced the jeep convoy to detour for 16 tense, sweltering hours along the Panj River, which forms part of the border with Afghanistan. The few metres between the road and the river were laid with landmines.

Still, Gasparini and Balli have caught the fieldwork bug, and are planning another expedition for next year. At the end of the nineteenth century, members of a small, isolated Italian population near Trieste emigrated in droves to South America — where they settled in different valleys in the Andes. Gasparini wants to compare these people's taste preferences and the acuity of their senses with those of their relatives in Italy. Unlike the Silk Road populations, the groups have the same genetic backgrounds but very different diets. Gasparini hopes that this will create a natural laboratory for exploring the interaction of diet and other environmental factors with interesting gene variants identified in the Silk Road samples.

Any discomforts of expedition science are overwhelmingly compensated for by the gains, says Gasparini. And if nothing else, the Marco Polo expedition taught him that boiled sheep's head is not his dish of choice. ■

Alison Abbott *is* Nature*'s senior European correspondent.*

MAN OF THE DESERT

BY QUIRIN SCHIERMEIER

Stefan Kröpelin stands near one of Chad's Ounianga lakes, an area that he helped to get listed as a World Heritage site.

Stefan Kröpelin has carved out a career where few dare to tread — in the heart of the Sahara.

Stefan Kröpelin's first research trip to the eastern Sahara was nearly his last. As a graduate student in January 1982, he travelled alone to Gilf Kebir — a remote plateau in the southwest corner of Egypt — to study sediment deposition in the steep canyons. Things didn't start well. On the first night, he found that his thin clothes and sleeping bag were no match for the freezing conditions. During the day, fierce sandstorms forced him to dig a pit for shelter and keep his eyes and mouth shut for most of the time. After a week, the driver who was scheduled to pick him up was nowhere in sight. Three days later, and running out of water, Kröpelin was just about to embark on a desperate 150-kilometre march to the nearest oasis when the jeep finally arrived.

Kröpelin, now a geologist and climate researcher at the University of Cologne in Germany, not only survived the ordeal but went on to become one of the most devoted Sahara explorers of our time. In dozens of expeditions to some of the farthest corners of the desert (see 'A career in the sand'), he has endured week-long sandstorms, suffered bouts of serious disease, including schistosomiasis, and faced heavily armed groups that roam the eastern Sahara.

But those decades of difficult field work have paid off for Kröpelin, who has made seminal discoveries about the climatic history of the Sahara that are challenging assumptions about the tipping points the world may face in a warmer future. At the same time, Kröpelin has worked to document the Sahara's cultural history and to preserve its heritage by lobbying to protect important scientific and cultural sites. "I've never ceased to be excited about the chance to investigate some of the least known parts of the desert," he says.

Siddiq Abd Algadir, president of the Sudanese Geologists' Union in Khartoum and a fellow student with Kröpelin in the 1980s, says that the German researcher has added immeasurably to Saharan science. "Much of what we now know about the geology, the environments and even the people in some of the most remote parts of the Sahara, we really owe to him and the expeditions he has led."

With his wiry build and boundless energy, Kröpelin seems much younger than his 60 years. The desks in his office are covered with parchment-coloured maps featuring vast expanses of desert. And his bookshelves sag with hundreds of volumes covering Africa's natural and cultural history, from early human evolution and migration to the trans-Saharan trade routes that carried gold and salt between West Africa and the Mediterranean.

In the 1960s, Kröpelin's sense of adventure was sparked by reading the exploits of nineteenth-century explorers such as Heinrich Barth and Gustav Nachtigal, who were among the first Europeans to document the cultures and environment of the Sahara. Two decades later, in graduate school at the Technical University of Berlin, that early curiosity about the area became a scholarly passion as he studied its physical geography and geology.

Similar in size to Australia or the United States, the Sahara is one of the most inhospitable and least explored spots on the globe. Throughout his work there, Kröpelin has tried to amass detailed knowledge of the scientific and cultural facets of his study areas before turning out results. Researchers who have travelled with Kröpelin describe him as a circumspect, hard-driving leader, equally at home in geomorphology, archaeology, human psychology and — when necessary — motor mechanics.

MAP QUEST

Over the past few years, Kröpelin's work has helped to reveal how the Sahara transformed from a savannah more than 5,000 years ago to the desert it is today. Researchers had previously thought that the transition happened abruptly — within little more than a century — when a cyclical shift in Earth's orbit reduced the amount of sunlight in the tropics and weakened the African monsoon. This idea was championed by Peter deMenocal of Columbia University in New York, who studied a core of deep-sea sediments from the tropical Atlantic ocean and saw a rapid rise in the amount of dust blowing off the continent at the time of climatic transition[1]. The sediment record agreed with models suggesting that a number of sensitive components of the planet — such as monsoon patterns, Amazon and Saharan vegetation, polar ice sheets and even the Atlantic Ocean circulation — will, under certain circumstances, flip from one stable state to another[2].

But Kröpelin wasn't convinced. The concept of an abrupt climate switch didn't mesh with his previous research on ancient settlements in the eastern Sahara[3]. "There is evidence from thousands of archaeological sites throughout the Sahara that prehistoric human settlements weren't abandoned within a few decades or so," he says. He was also piqued that deMenocal reached his conclusion without ever setting foot in the desert, and used a single marine record to make generalizations about the entire Sahara. "The idea of catastrophically fast climate change is untenable — it can only come from someone who doesn't know the Sahara," says Kröpelin.

From his past reconnaissance work, Kröpelin knew one place in the Sahara with a climate record potentially long enough to test his ideas — a series of permanent lakes in the remote Ounianga basin in northeastern Chad, a practically rainless and notoriously insecure region bordering Libya and Sudan. French military geographers had explored the lakes in the early twentieth century but the area had since been forgotten by most foreign researchers. The lakes are fed by ancient groundwater, and their depths are layered with sediments containing climatic indicators such as pollen. Kröpelin suspected that one of the largest lakes, Lake Yoa, held a sedimentary cache thick enough to reconstruct the long-term climatic history of the region.

Working in Ounianga has its challenges. On a reconnaissance mission in 1999, Kröpelin had to persuade a superstitious local population that drilling a hole in the lake bottom would not drain the water or release the jinns and demons they feared might dwell in the ground.

After extensive preparations, he returned to the remote lake in 2003 and 2004, and pulled up a continuous 8-metre-long core of sediment from its bottom — a climate record reaching back 6,000 years. Kröpelin's analysis revealed that it took some 3,000 years — from 5,600 to 2,700 BC — for the fully vegetated savannah there to transform into a barren desert[4].

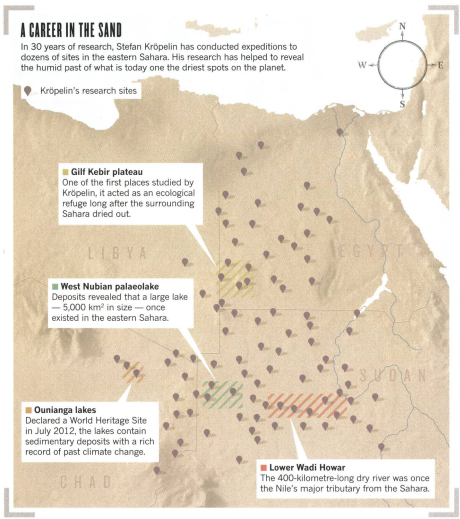
THE GREAT DRYING

At first, some researchers questioned the dating of the core and whether it represented conditions across a broad region. But Kröpelin answered his critics[5], and his ideas eventually gained acceptance. Today, deMenocal maintains that the western Sahara dried out quickly, but he acknowledges that the eastern part around Lake Yoa took much longer. "Stefan knows the Sahara better than any other living scientist, and the climate archive he has uncovered, while not perfect, is uniquely well-dated," he says.

The debate has intrigued climate modellers interested in potential tipping points in the Earth system, in which small perturbations cause abrupt and major changes. Kröpelin's findings imply that the most recent savannah-to-desert transition does not fit this pattern — although he leaves open the possibility of earlier tipping points in the Saharan climate.

Climate researchers are therefore eager for Kröpelin to report his analysis of a 16-metre-long core from Lake Yoa that he drilled in 2010, which he is now readying for publication. "The published record is fascinating but rather

S. KRÖPELIN/UNIV. COLOGNE


A CAREER IN THE SAND
In 30 years of research, Stefan Kröpelin has conducted expeditions to dozens of sites in the eastern Sahara. His research has helped to reveal the humid past of what is today one the driest spots on the planet.
Kröpelin's research sites
N
W
E
S
Gilf Kebir plateau
One of the first places studied by Kröpelin, it acted as an ecological refuge long after the surrounding Sahara dried out.
West Nubian palaeolake
Deposits revealed that a large lake — 5,000 km² in size — once existed in the eastern Sahara.
Ounianga lakes
Declared a World Heritage Site in July 2012, the lakes contain sedimentary deposits with a rich record of past climate change.
Lower Wadi Howar
The 400-kilometre-long dry river was once the Nile's major tributary from the Sahara.
LIBYA
EGYPT
SUDAN
CHAD

SOURCE: S. KRÖPELIN

short," says Martin Claussen, a climate modeller at the Max Planck Institute for Meteorology in Hamburg, Germany, who is exploring ways to predict fast climate–vegetation transitions[6]. "The complete 12,000-year record should help us get a better sense of how rapidly the system might be capable of switching from one state to another and what the transition period may look like."

The results from Lake Yoa crown a long list of discoveries that Kröpelin has made in the region. In one of his earliest major finds, Kröpelin established that the dry valley known as Wadi Howar, which sits in an extremely arid part of northern Sudan, was once one of Africa's largest rivers and a tributary to the Nile[7]. This extinct river flowed from about 9,500–4,500 years ago and supported a rich savannah that was home to a host of animals, including antelopes, giraffes, zebras and elephants.

Kröpelin's extended trips through the Sahara, which typically last two months, have sometimes placed him in difficult situations. During one expedition to northern Sudan in 1995, he was chased for several days across the desert by a truck. When the pursuers caught him, they turned out to be Sudanese forces who thought they were following outlaws smuggling weapons or people. The soldiers were nearly out of fuel from their chase, so Kröpelin gave them a share of the expedition's diesel supplies and they enjoyed a friendly meal together. The excavations carried out on that trip revealed how prehistoric humans settled there during the Sahara's humid phase and then disappeared as the climate dried out[3].

Another experience in 2005 was more troubling. One night, the sound of heavy gunfire terrified a team that Kröpelin had led to Erdi Ma, an uninhabited and largely unexplored tableland in the northwest corner of Chad. The next morning they learned that Sudanese military forces had massacred 20 Darfur rebels not far from the expedition's campsite.

Sometimes, nature is the biggest enemy. On the way to Lake Yoa in April 2010, Kröpelin and his team got caught in a dust storm fiercer than any he had seen in 30 years. For more than a week visibility dropped to two metres. Under non-stop bombardment, the researchers' bodies and vehicles became electrically charged and each touch caused painful electric shocks. Refilling the trucks from metal gas canisters was Russian roulette, as large sparks shot around the filler necks. "Back home, it took weeks until I could open my car again without inhibition," Kröpelin says.

Through his expeditions, Kröpelin has developed relationships with many of the regional governments, and he has helped to advise officials in Egypt, Libya and Chad on the impact of mining, oil drilling and tourism. He has campaigned for and won protection for two environmentally and culturally unique regions that he explored — Wadi Howar and Gilf Kebir.

For the past 13 years, Kröpelin has also lobbied to protect the Ounianga lakes — an effort that involved hundreds of meetings with ministers in Chad and international agencies and officials. That work paid off in July, when the region was declared a world heritage site by UNESCO (the United Nations Educational, Scientific and Cultural Organization).

Researchers wonder what will happen when Kröpelin retires, leaving a major hole in Saharan studies. "He has taken risks few researchers would be happy to take any more," says deMenocal.

The risks in the region have grown lately because of rebel activity and recurring armed conflict in places including Darfur, Libya, Mali and Chad. But despite some anxious moments over the years, Kröpelin considers that the risk of getting robbed, kidnapped or killed is smaller in the Sahara than in sections of London or New York City.

"In the desert there is an air of fellowship with whoever crosses your way," he says. "Encounters with nomads, rebels and more shady folks can be scary, but I've always managed, somehow, to talk us out of even the most delicate incidents."

For now, Kröpelin shows no signs of slowing down. He plans to return to the Sahara in November for a celebration of the UNESCO decision and to do some geological field work. Next year, he will search for traces of prehistoric writing and trading, and will explore a mysterious crater in Jebel Uweinat, the tallest mountain in the eastern Sahara. Gone are the days of foolhardy solo trips, but not the deep love for the land. "Under the vast desert skies, your whole existence takes on a different meaning," he says. "I'd like to return for as long as I can." ■

Quirin Schiermeier *is a reporter with* Nature *in Munich.*

1. deMenocal, P. *et al. Quat. Sci. Rev.* **19,** 347–361 (2000).
2. Lenton, T. M. *et al. Proc. Natl Acad. Sci.* **105,** 1786–1793 (2008).
3. Kuper, R. & Kröpelin, S. *Science* **313,** 803–807 (2006).
4. Kröpelin, S. *et al. Science* **320,** 765–768 (2008).
5. Kröpelin S., Verschuren D. & Lézine A.-M. *Science* **322,** 1326c (2008).
6. Bathiany, S., Claussen, M. & Fraedrich, K. *Clim. Dyn.* **38,** 1775–1790 (2012).
7. Pachur, H.-J. & Kröpelin, S. *Science* **237,** 298–300 (1987).

COMMENT

SPACE Did supermassive black holes drive the evolution of life itself? **p.278**

ART New York exhibition explores the surreal side of technology **p.279**

TECHNOLOGY Talking to a creator of therapeutic robots **p.280**

BIODIVERSITY Conservation triage is sometimes necessary **p.281**

M. FUNK/PHOTOGRAPHER'S CHOICE/GETTY

Uses such as lighting, heating and cooling in buildings account for almost 40% of the energy consumed in the United States.

How to build a low-energy future

Advanced construction technologies promise huge energy savings, says **Philip Farese**. Investment is needed to bring them to market and to encourage their use.

In 2010, the United States used 98 quadrillion British thermal units (quads) of energy — about 20% of the world's total — on everything from transport to entertainment. A surprisingly large amount — 40 quads — went into buildings for uses such as space heating, cooling, lighting and computing. Assuming that business continues as usual, by 2030, the nation will require 43 quads of energy for buildings each year[1].

There is huge room for improvement without sacrificing an ounce of comfort. Much more should be invested in developing energy-saving technologies, in bringing them to market and in encouraging their use.

I, along with colleagues at the National Renewable Energy Laboratory (NREL) in Golden, Colorado, and at the US Department of Energy (DOE), have reviewed more than 300 publications and sought the advice of 65 experts to determine how much US energy demand for buildings might be cut. We find[2] that cost-effective technologies available now could reduce energy use in buildings by up to 30% by 2030, which is in line with others' findings[3].

Add in the affordable emerging technologies that are expected to be available in the next 5 years or so, and the potential reduction reaches 55%. Including all ▶

technologies that are in the early stages of research and development (R&D) brings this cut in energy use to an astonishing 80% (see 'Going down').

Some researchers have pointed out that efficiency improvements can lead to an increase in demand[3]. Such 'rebound' effects are real, but are expected to soak up no more than 15% of energy savings[3]. Some people might heat their homes a bit more in the winter, but very few will wash more clothes or cook more meals. Other researchers argue[4] that the money saved through energy efficiency will be spent by consumers on energy-intensive luxuries, such as long-distance flights. But consumers are likely to spend in line with their usual budget. For every dollar spent in the United States, about 8% goes back to energy costs, including the energy used to manufacture, sell and transport goods — an indirect rebound effect.

Even accounting for these effects, energy savings should still exceed 20% from currently available technologies and 45% from emerging technologies[2]. The 80% reduction as a result of advanced technologies becomes 65% when rebound effects are included, meaning that the projected 43 quads of demand could be met using some 14 quads of energy. The national energy bill could be reduced by as much as US$275 billion a year[2] if just a fraction of that amount was invested strategically.

According to my analysis, energy-use intensity — the amount of energy used per unit area of building space per person — actually rose in the United States by 10% from 1980 to 2010. Because of this, some say that efficiency measures serve to increase energy use. But this rise results from an increase in demand for modern devices such as air-conditioning units and computers. Heating (of both water and space) and refrigeration have also seen a small rise in demand despite significant improvements in equipment efficiency — but by my calculation, their energy-use intensity has fallen by 43%. Efficiency will need to be pursued aggressively in all areas, using approaches that range from increasing R&D investment to maintaining appliance standards.

HEAT AND LIGHT

For centuries, most of the energy used to heat buildings came from burning fuels, which can provide no more than one unit of heat for each unit of energy. But many energy-efficiency technologies are in the pipeline, and a few are particularly revolutionary.

Heat pumps could be much more efficient. They currently exchange at least three units of heat from inside and outside buildings per unit of electricity. When the thermodynamic effect of producing electricity is accounted for, the pumps become about as efficient as the best available gas furnaces. More than 11 million homes in southern and temperate climates in the United States already use them. Further savings could be made by improving the technologies used and by lowering the installation cost of pumps that extract heat from the ground. These can swap an impressive 8 units of heat for every unit of electricity. For water heating, we should encourage the adoption of electric and gas-fired pumps and solar heaters. Devices are now available that more than double the efficiency, at half the lifetime cost.

Meanwhile, liquid-desiccant technologies are starting to shift how people think about cooling and de-humidification. Current technology relies on the compression and expansion of vapour to drive heat out of a building. To remove moisture from the air, these often over-cool the air, then heat it up again. Solutions of brine, however, are much more efficient at extracting humidity. These systems use 30–50% less energy and are already commercially available from firms such as Advantix Systems in Sunrise, Florida, of which I am a vice-president. Systems that combine desiccant and evaporative cooling are on the horizon, and promise savings of 60–80%.

For lighting, simply increasing the use of efficient technologies such as compact fluorescent bulbs could lessen the energy used by 35%. The biggest opportunity lies with light-emitting diodes (LEDs), which could push energy use down by as much as 90%. The DOE awarded Amsterdam-based electronics firm Philips $10 million for producing an LED lamp that uses 85% less energy than an incandescent bulb, lasts roughly 25 times longer and provides the same quality of light. The bulb costs $50, but saves the consumer $200 in energy costs over its lifetime, compared with incandescent bulbs. The high purchase price is a sticking point, but manufacturers and the DOE plan to get the cost down to just $5 in the next 5–10 years.

"'Rebound' effects are are expected to soak up no more than 15% of energy savings."

Simple improvements to building insulation can also make a difference. For example, highly insulating windows currently cost about $1,000 more than standard windows for a typical home, with that difference earned back through energy savings in 5–10 years. But the profit differential isn't large enough for distributors to justify the time needed to convince customers to buy the product. One solution could be to create web tools or mobile apps that easily showcase the advantages of efficient products.

Savings of up to 20% or more are available through changes to maintenance and operations procedures. For example, 'learning thermostats' in homes assess the daily schedule and adjust the temperatures of

SOURCE: REF. 2

E. DUNAND/AFP/GETTY

A heating and cooling system based on geothermal energy at a hospital in Greensburg, Kansas.

the rooms accordingly, and 'tune ups' on commercial buildings improve how they are run. Cars use control systems that are much more intelligent than those in our buildings.

INVEST TO SAVE

The barriers to 80% energy savings in buildings are well known. Consumers find it hard to evaluate the risks and benefits of technologies; investors may not benefit directly (why would a landlord invest in LED bulbs when tenants pay their own electricity bills?); and few of the latest technologies have been taken up by vendors. These small problems add up to big obstacles. But there are solutions.

Energy standards for US appliances were put in place between 1987 and 2010, for example, leading to savings of 3 quads in 2010. These are forecast to save a staggering 158 quads by 2070 (ref. 6). The DOE's Appliance and Equipment Standards programme is required by law to update appliance standards, but that takes effort. The most recent refrigerator standards, released in 2011, were developed after years of work, culminating in a 1,000-page report. Despite offering a return on investment of about a thousand to one, the rate of issuing standards has waxed and waned over time, and the programme has not always received the funding it needs. It is now on track, but no one knows for how long.

Building codes that were put in place in the 1970s have reduced energy use in newly constructed buildings by 30–40%, leading to a saving of 0.45 quads per year[7]. Codes developed in 2010 will soon be in place that should save an extra 1.25 quads per year by 2030 — a huge improvement — and others in development could double those savings[7]. However, every state will need to adopt these codes individually, and many will need help with compliance.

Labels can be powerful, too. Since the DOE and the US Environmental Protection Agency introduced the ENERGY STAR programme in 1992 to test and label the efficiency of household equipment, energy use has fallen by about 1.6 quads per year[8]. Like appliance standards, this programme needs to be updated regularly. As new products raise the bar, old products need to lose their status. A recently introduced 'most efficient' ENERGY STAR label is a step in the right direction.

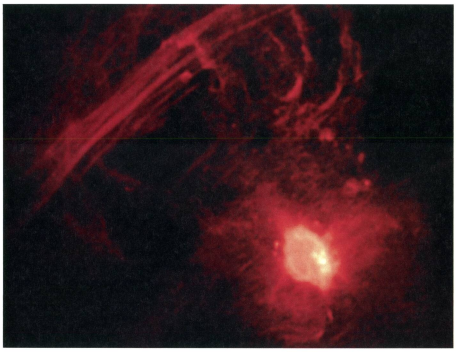
Further investment in R&D could make a difference. A 2001 analysis[9] of selected DOE-funded research projects showed net energy cost savings of around $30 billion from an investment of about $0.4 billion in energy-efficiency R&D: a return of 75 to 1.

Despite such rewards, the energy sector is hugely underfunded. Total R&D investments from US public and private sources come to just 0.4% of the amount spent on energy bills (about $1.5 billion per year in buildings-related R&D). This percentage is significantly lower than that spent in the pharmaceutical (20.5%) and aerospace and defence (11.5%) sectors[10]. Assuming the same return of 75 to 1, $40 billion would need to be invested to generate a cumulative $3 trillion of savings. Even if only $3 billion were invested per year until 2030, that would still be only about 1% of the money spent on buildings-related energy.

Strategic investment offers the opportunity to reduce the US annual energy bill by hundreds of billions of dollars. Achieving an energy-efficient future is a cause that can benefit all of us. ■

Philip Farese *is vice-president of sales support at Advantix Systems in Sunrise, Florida, the leading manufacturer of liquid-desiccant air conditioners. He was a senior energy analyst at the US National Renewable Energy Laboratory until early August 2012.*
e-mail: philip.farese@advantixsystems.com

1. US Energy Information Administration *Annual Energy Outlook 2012* (Department of Energy, 2012).
2. Farese, P., Gelman, R. & Hendron, R. *A Tool to Prioritize Energy Efficiency Investments* (National Renewable Energy Laboratory, in the press).
3. *Real Prospects for Energy Efficiency in the United States* (National Academies Press, 2009).
4. Jenkins, J., Nordhaus, T. & Shellenberger, M. *Energy Emergence: Rebound & Backfire as Emergent Phenomena* (Breakthrough Institute, 2011).
5. Sorrell, S., Dimitropoulos, J. & Sommerville, M. *Energy Policy* **37**, 1356–1371(2009).
6. Meyers, S., Williams, A. & Chan, P. *Energy and Economic Impacts of U.S. Federal Energy and Water Conservation Standards Adopted From 1987 Through 2010* (Lawrence Berkeley National Laboratory, 2012).
7. Belzer, D. & Cort, K. A. *Projected Impacts of the Building Energy Codes Program, 2012–2030* (Pacific Northwest National Laboratory, 2012).
8. Sanchez, M., Brown, R. E., Homan, G. K. & Webber, C. A. *2008 Status Report Savings Estimates for the ENERGY STAR Voluntary Labeling Program* (Lawrence Berkeley National Laboratory, 2007).
9. National Research Council *Energy Research at DOE: Was It Worth It? Energy Efficiency and Fossil Energy Research 1978 to 2000* (National Academies Press, 2001).
10. Holliday, C. *et al. Catalyzing American Ingenuity* (American Innovation Energy Council, 2010).



NRAO/AUI/NSF/F. ZADEH ET AL.

Radiowave images show filamentary structures near the black hole at the centre of our Galaxy.

COSMOLOGY

Enigma at the heart of the Universe

Mario Livio enjoys a tour of modern astronomy that speculates on how black holes drive cosmic evolution.

"The road of life twists and turns and no two directions are ever the same. Yet our lessons come from the journey, not the destination." So wrote US novelist Don Williams Jr. Caleb Scharf's engaging *Gravity's Engines* demonstrates the truth in this statement.

Scharf speculates that black holes rule everything in the cosmic landscape — from the large-scale structure of the Universe to life. Using rich language and a brilliant command of metaphor, he takes on some of the most intricate topics in theoretical and observational astronomical research. He weaves a wonderfully detailed tapestry of what modern astronomy is all about, from the complexities of cosmic microwave background studies to the X-ray mapping of galaxy clusters.

Scharf begins with the basics: an excellent overview of the history of astronomy and astrophysics, and illuminating examples to explain how the complete cosmic picture is deciphered. His discussion of how gravity shapes the Universe and determines its evolution is admirably clear, and covers the early history of X-ray and radio astronomy, as well as some of the perplexing objects (such as compact X-ray sources) discovered using new observational techniques.

Gravity's Engines: How Bubble-Blowing Black Holes Rule Galaxies, Stars, and Life in the Cosmos
CALEB SCHARF
Scientific American: 2012. 272 pp. $26/£18.99

He handles the complex physical processes that occur in the vicinity of black holes with aplomb and attention to detail. And he gets to the crux of the feedback process between the growth of mass-accreting black holes and the evolution of stellar bulges in their host galaxies.

I feel that Scharf begins to overstate his argument in his attempts to solidify the case for supermassive black holes as "regulators" of star formation in galaxies. There is no doubt that central black holes and the stellar bulges that surround them (when they exist) co-evolve. But the outer disks of spiral galaxies such as Andromeda are hardly affected at all by the presence of supermassive black holes.

I have quibbles, too, with the passages in which Scharf attempts to support his argument that black holes are the main engines driving everything from re-ionization and cosmic star formation to galactic evolution and the emergence of life. For instance, he argues that stellar-mass black holes played a crucial part in the cosmic phase transition associated with the re-ionization of the intergalactic medium. But the jury is still out on this particular process. The main agent of re-ionization may be ultraviolet photons from galaxies — a point that looks likely to be verified soon through observation.

Scharf also uses the story of our own Galaxy, the Milky Way, and statistical information collected through the citizen-science project Galaxy Zoo, to depict the "duty cycles" of activity and calm in the process of accretion onto supermassive black holes. He appeals to these to circumvent the apparent discrepancy between the idea of black holes as regulators and the fact that the Milky Way's central black hole is currently inactive.

All this culminates with his intriguing statement that "the entire pathway leading to you and me would be different or even nonexistent without the coevolution of galaxies with supermassive black holes and the extraordinary regulation they perform". Scharf admits that many steps remain uncertain and that numerous questions have yet to be answered. But he proposes that because the cosmic and galactic environments leading to the rise of complexity and life are part of black holes' galactic evolution, it is reasonable to ask what special things link us directly to that history.

However, I feel less certain than Scharf about the answer. He explains that the tight correlation between the masses of supermassive black holes and of stellar bulges at galaxies' centres reveals a co-evolution. It is equally certain that feedback from supermassive black holes had an important role in the ensuing star-formation history in the bulges of galaxies. But was this the key factor in determining whether life-bearing planets should exist or not? I doubt it. Still, the idea makes for a very interesting journey. ■

Mario Livio *is an astrophysicist at the Space Telescope Science Institute in Baltimore, Maryland. His blog, A Curious Mind, explores science, art and the links between them. His upcoming book is* Brilliant Blunders, *to be published in May 2013. e-mail: mlivio@stsci.edu*

TECHNOLOGY

Pre-digital dreams

Josie Glausiusz strolls through an evanescent 'cabinet of wonders' exploring the surreal side of technology.

I flung myself onto the floor cushions inside Stan VanDerBeek's *Movie-Drome* and gazed at the flickering images projected on its curved ceiling. An attack of insomnia the night before gave me a good excuse to lie down and relax as pictures spun by: a dancing skeleton, a fork, a statue, a fried egg, a terrified face, the Statue of Liberty, the Kremlin, the words 'WHO WHAT HOW'. It seemed as if I was inhabiting a frenetic early version of the Internet — which, in a way, I was.

VanDerBeek, an experimental film-maker, constructed the original 9-metre-high movie-drome inside a grain silo in 1963. He foresaw a global network of such pods as "hubs for the distribution of knowledge via the universal language of the information age", connected by a network of satellites, television and telephones. He died in 1984, having never realized his dream. The reconstructed pod — a piece of this World Wide Web prototype — forms the centrepiece of *Ghosts in the Machine*, an exhibition at the New Museum of Contemporary Art in New York.

The exhibition is conceived as a "cabinet of wonders" encapsulating a prehistory of the digital age. It is designed to explore "the dreams and nightmares of technology" as expressed by more than 70 artists, writers and other visionaries. The eclectic jumble of installations conveys both our dread and our embrace of technology, with its power to transform life and death.

The most terrifying is a reconstruction — commissioned for a 1975 exhibition — of the *Harrow*. This fictional execution machine appears in Franz Kafka's short story *In the Penal Colony* (1919). In Kafka's tale, a person accused of an unspecified crime is strapped to a bed while an array of long needles carves the purported crime into their flesh.

With its cogs and gears, dangling wires and sharp needles suspended over a coffin-like bed, the wooden contraption evokes the terror that must face every prisoner confronted with today's electroshock devices and other torture technologies. Incongruously, the *Harrow* shares a space with Jeff Koons's vacuum cleaners, the *New Hoover Convertible Doubledecker*, encased in Plexiglass and lit with fluorescent bulbs. Koons's piece is perhaps there to evoke the torment of domestic drudgery.

Another misconceived machine in the show, created by psychoanalyst Wilhelm Reich in 1940, is the Orgone Energy Accumulator, a wooden cubicle lined with sheet metal and steel wool. A follower of Freud, Reich believed that a concentration of "orgone", which he conceived of as a primordial form of energy, would increase "orgastic potency". This would, in turn, relieve pent-up psychic and physical tensions, promoting general health and vitality, and even playing a part in curing cancer. Reich offered the box for testing to Albert Einstein, who pronounced it a dud.

Among those treated with orgone therapy was British computer-science pioneer Alan Turing, who was forced into treatment to "cure" his homosexuality. His proposed Turing Machine would simulate the behaviour of any other machine, including itself. Two years after being prosecuted for gross indecency in 1952, Turing committed suicide, possibly by eating a cyanide-laced apple.

Henrik Oleson's *Some Illustrations to the Life of Alan Turing* (2008) appears in the same room as Reich's accumulator. This photo-collage series representing the dualities of Turing's life — public and private, mind and body, human and machine — depict Turing's face obscured by, among other objects, a large screw, a grid of zeroes and ones, and a half-eaten apple.

The machine as medicine shows up again in a piece of art constructed by Emery Blagdon, a sometime vagabond and tinkerer who once built a working tractor from the ground up. Born in 1907, Blagdon created intricately woven wire-and-bead chandelier-like mobiles, which he called "healing machines", and hung them in an outhouse on his Nebraska farm. He re-routed the power lines from his house into the shed, hoping to harness the electricity's healing powers. However, Blagdon died of cancer in 1986, the energy from his machines having failed to heal him.

There is much to admire in *Ghosts in the Machine*. Thomas Bayrle's 1989 *Madonna Mercedes*, for instance, is a mother-and-child collage that, on close inspection, is made up entirely of drawings of Mercedes-Benz automobiles. Günther Uecker's *New York Dancer IV* (1965), a nail-studded phallic sculpture 2 metres tall and powered by an electric motor, is more mystifying.

Mystification, however, can also be a thing of beauty and delight, as evinced by Gianni Colombo's *Elastic Space*, a three-dimensional grid made of illuminated, moving rubber bands. Designed to "catalyze variable perceptual experiences", the piece is mesmerizing and soothing.

I stood in the quiet, darkened room watching the glowing rubber bands slowly moving back and forth, up and down. I thought perhaps I would stay for a while. I could contemplate infinity. And maybe go to sleep. ■

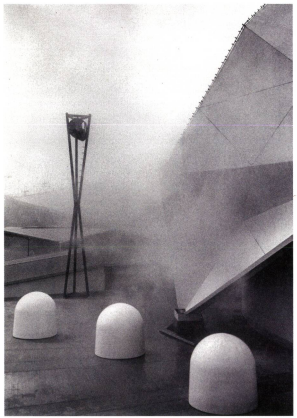
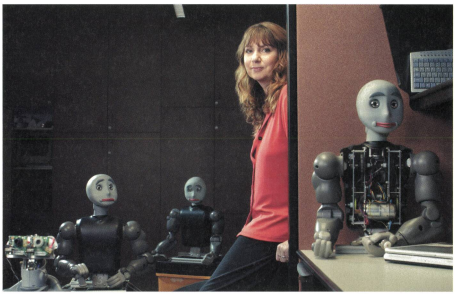


PHOTO: SHUNK-KENDER. 1970/® ROY LICHTENSTEIN FOUNDATION

Robert Breer's *Floats* (1970) move imperceptibly about the gallery.

Ghosts in the Machine
New Museum of Contemporary Art, New York City. Until 30 September 2012.

Josie Glausiusz *is a journalist based in New York City.*
e-mail: josiegz@gmail.com



ERIC OGDEN/CORBIS OUTLINE

Q&A Maja Matarić
The social roboticist

Maja Matarić, a computer scientist and neuroscientist at the University of Southern California in Los Angeles, makes robots that assist people with disabilities, children with autism and elderly people — a phenomenon explored in the film Robot and Frank (2012). *On the eve of its release, she talks about the future of socially assistive machines.*

How does *Robot and Frank* portray the human–robot interface?

The film asks whether people can relate to machines, and how robots can bridge the gap between what ageing people need and what their families can provide. In it, an elderly jewel thief convinces his caretaker robot to help him pull off a heist.

Did the film use a real robot?

The film-makers make a human in a robot suit look like a robot trying to look human. The robot's capabilities are necessarily beyond the current state-of-the-art: it serves as butler, nurse, chef and maid. Robots that manipulate objects in the real world are harder to engineer, and because these tasks require coordination and strength, safety is a major concern.

Are you designing those kinds of robots?

There's a whole realm of need that doesn't demand any physical work. Robots can help people to cope with depression, isolation, exercise and even social disorders such as autism. Most robotics is about getting the robot to do the right thing, whereas we make robots that get the person to do the right thing. Our socially assistive robots provide coaching, encouragement and companionship. We would never want to coerce people; rather, we aim to influence their behaviour. Someone who has had a stroke may need to spend hours each day working to regain function in an arm. A robot can use persuasion, motivational techniques and social manipulation to help the patient to persist with that. One student in my lab is studying how a robot might use the human quality of reciprocity. By helping a person with a task, such as shelving books, it would make it more likely that the person would adhere to the robot's requests, such as to do more exercise.

Can robots help with diagnosing autism?

With about one in every 100 children affected, early diagnosis is badly needed. We are designing a robot for diagnostic use in paediatric waiting rooms: as kids play with it, it could flag those with autism-like symptoms.

What about treatment of autism?

We have demonstrated that child-sized humanoid robots can encourage some children with autism to be more verbal and empathetic. The robots' life-like appearance and responsive behaviour seem to stimulate children to play with them and express empathy; when our robot didn't obey a command, one child said, "Now I know how my teacher feels." We're using a robot to model an autism behavioural therapy that aims to help a child gradually improve on a task. It will use verbal feedback and encouragement while playing games to encourage the child to imitate its poses and gestures. Ultimately, we hope to see the results of this research in affordable products for home use.

NATURE.COM
For more on the ethics of robotics, visit:
go.nature.com/lwx6ow

How can robots motivate people?

In a study we used Bandit, a child-sized humanoid robot on wheels, as an exercise coach. A person was seated across from Bandit and they played 'chair aerobics'. Sometimes Bandit led the exercises, and sometimes the user taught exercises to Bandit — many people love to be in charge some of the time. In another game, Bandit gave ever-longer exercise sequences to test memory and stamina. Through the use of computer vision, Bandit was able to detect the person's behaviour and performance, and respond appropriately by providing encouragement, increasing the levels of challenge and making humorous comments. We found that users strongly preferred Bandit to a computer simulation with the same games and behaviours, rating the robot as knowledgeable and caring. It effectively motivated people to exercise consistently.

Will caretaking robots take jobs from people?

Thirty years from now, there will be about as many old people as children and working-aged people in the United States. The population demographics are already creating a gap in care. Even if we expanded the immigration rate, we couldn't afford to fully train and pay everyone to be caregivers for the entire elderly population. Technology, including robotics, will have to be part of the solution.

Do assistive robots pose ethical issues?

I am concerned about claims that salespeople might make about these machines' capabilities. We don't want snake-oil promises of health benefits that aren't clinically validated. Other issues could arise from companies failing to advertise necessary software upgrades or implement recalls. And, as with people, when you put robots in the 'wild', their behaviour is not entirely predictable. If robots are continually learning and adapting to their users, it will be interesting to see the relationship dynamics that emerge between the two.

How can long-term robot–human relationships be managed?

We don't want attachments between humans and their robots to become dysfunctional. As with human–human relationships, it is best that people do not become overly dependent, submissive or abusive. Some ask what will happen if a well-loved robot breaks down. Its memory will be backed up and, for a price, an identical body can be built for its data to live in. The real question is, what is the warranty? ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Keep politics out of academia in Israel

Israel's thriving academic establishment is being ripped apart by the Israeli government's apparent drive to use academics for political ends.

To bolster its agenda of developing settlements in the West Bank occupied territories, the Israeli military's Council for Higher Education in Judea and Samaria has upgraded a small college there, formerly known as Ariel College, to university status. Millions of dollars are to be poured into the university's budget to attract researchers, circumventing Israel's usual channels for academic funding.

In response, more than 1,100 senior faculty members from all over Israel — over one-quarter of the country's total — have signed a petition rejecting the upgrade (see go.nature.com/dqte9c). The planned upgrade is also opposed by the presidents of all seven research universities in Israel and by the influential planning and budgeting committee of the state's Council for Higher Education.

The legitimacy of building settlements on occupied land is disputed in Israel and deplored worldwide. Ever since the occupation of the West Bank 45 years ago, Israeli academia has largely remained outside the political struggle. But Ariel University, which is closed to Palestinians, creates a dilemma for academics there: participate and thereby endorse the settlements, or boycott and be branded as the enemy. Such political meddling can only ruin Israel's academic structure in the long run.

Nir Gov *Weizmann Institute of Science, Rehovot, Israel.*
nir.gov@weizmann.ac.il

Turkey must build on science reforms

As chair of the Committee on Freedom and Responsibility in the Conduct of Science at the International Council for Science (ICSU), I was alarmed to read of the arrest in Turkey in June of Kemal Gürüz (http://doi.org/h47), twice a victim of the country's apparent crackdown on academic freedom (*Nature* **487,** 19–20; 2012). The committee monitors such cases and calls for greater protection of the human rights of individual scientists.

Gürüz, as a former president of Turkey's Council of Higher Education, has helped to raise education standards significantly in the country. He has also facilitated higher-education opportunities abroad for hundreds of Turkish students — particularly in the United States, where he was a Fulbright scholar and, later, chairman of the Turkish Fulbright Commission. His rearrest seems to relate to his attempted modernization of the education system, including provisions to educate Turkish students to become scientists.

Turkey's scientific contribution will depend on building up its research structure and continuing the reforms initiated by influential academics such as Gürüz.

Bengt Gustafsson *ICSU, Paris, France.*
bengt.gustafsson@fysast.uu.se

Time to accept conservation triage

Like the troubled California condors *Gymnogyps californianus* (*Nature* **486,** 451; 2012), more than 80% of endangered US species are imperilled by threats that cannot be eliminated, only managed. These species are "conservation reliant" (J. M. Scott *et al. Conserv. Lett.* **3,** 91–97; 2010).

For example, the endangered Kirtland's warbler (*Setophaga kirtlandii*) has exceeded its recovery goals as a result of control of the brown-headed cowbird (*Molothrus ater*), which lays its eggs in warbler nests, and maintenance of warbler habitat through vegetation management. These conservation actions will be required in perpetuity if the species is to avoid extinction. But funding of such management is inadequate for all but the most iconic species, such as the condor.

Society must commit to increased investment in conservation intervention, which will take greater political motivation and broader conservation partnerships than at present. We must begin to prioritize conservation efforts and accept that conservation triage may sometimes be necessary.

John A. Wiens *PRBO Conservation Science, Corvallis, Oregon, USA. jwiens@prbo.org*
Dale D. Goble, J. Michael Scott *University of Idaho, Moscow, Idaho, USA.*

Call for clear policy on deceased authors

William Laurance and his colleagues (*Nature* http://doi.org/h5f; 2012) identify four of their paper's authors as deceased and confirm that all had previously verified that they wanted to be co-authors. Without such indication, there is a risk that deceased authors might be included on questionable grounds — for example, as little more than a dedication.

Increasingly, authors' contributions to papers have to be explicitly justified. Although opinions of what constitutes enough to warrant authorship can vary between individuals and journals, most would endorse your journal's policy that "all the listed authors [must] have agreed all of the contents" of the final version of the submitted paper (see go.nature.com/qnanep).

Because this criterion may not be fulfilled for a deceased author who has genuinely contributed to the work, journals should not only denote authors who died before submission but also specify the date of their death and the contribution they made (see, for example, go.nature.com/pymlxb and go.nature.com/jxqr8n). The author list should not include unwarranted names of the deceased, for whom a dedication can be included in the acknowledgements.

Vincent Nijman *Oxford Brookes University, UK.*
vnijman@brookes.ac.uk

How to allay fears over nanomaterials

Regulating the use and engineering of nanomaterials is problematic owing to scientific uncertainty and controversy over their safety. To pave the way for regulatory approval and to build consumer confidence, the European Commission's Joint Research Centre and the European Academies Science Advisory Council have identified factors to help assess the benefits and risks of nanomaterials to health (see go.nature.com/utmi3d).

Our analysis indicates that formulation of a coherent public policy will depend on scientists closing knowledge gaps in safety research, on gathering more data to connect science and regulation, and on training graduate students in nanotechnology research. Policies will need to be flexible to accommodate fresh discoveries in this rapidly advancing technology.

More dialogue on nanomaterials is needed among private and public research groups, regulators and the public.

Genetic modification of crops has taught us that consumer acceptance as well as regulatory approval are needed for successful implementation of new bioscience technologies. To limit alarmist media assertions about nanomaterials, scientists have a responsibility to provide accessible and accurate information for the public.

Robin Fears *European Academies Science Advisory Council, Halle, Germany.*
Peter Gehr *University of Bern, Switzerland. gehr@ana.unibe.ch*
Elke Anklam *Institute for Health and Consumer Protection, EC Joint Research Centre, Ispra, Italy.*

STEP UP

the pace of your research with

SOLID PHASE CO-TRANSFECTION

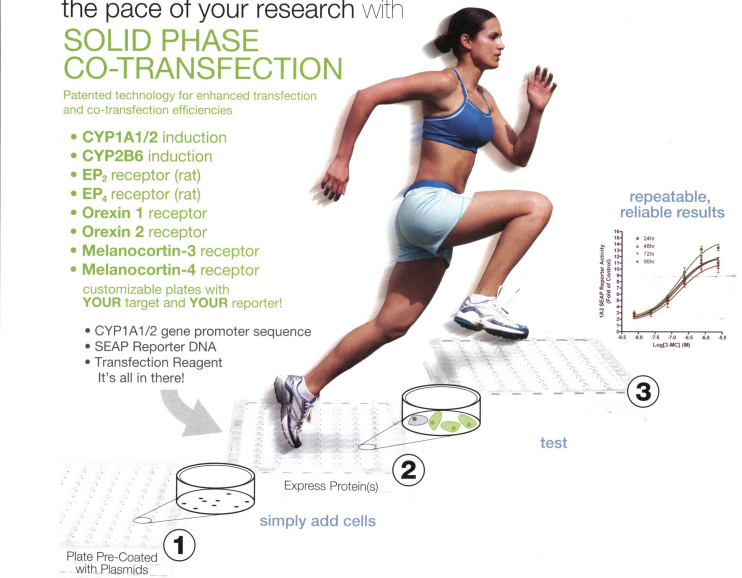
Patented technology for enhanced transfection and co-transfection efficiencies

- **CYP1A1/2** induction
- **CYP2B6** induction
- **EP$_2$** receptor (rat)
- **EP$_4$** receptor (rat)
- **Orexin 1** receptor
- **Orexin 2** receptor
- **Melanocortin-3** receptor
- **Melanocortin-4** receptor

customizable plates with **YOUR** target and **YOUR** reporter!

- CYP1A1/2 gene promoter sequence
- SEAP Reporter DNA
- Transfection Reagent

It's all in there!



STEP* (Surface Transfection and Expression Protocol)

Convenient
Ready to assay or store up to one year

Low Cost
Use routine cell lines, no need to maintain multiple stable cell lines

Reliable and Robust
Consistently high transfection and co-transfection efficiency

Highly Reproducible
Low variability well-to-well, plate-to-plate, and batch-to-batch

Allows Side-By-Side Comparison
Eliminate variations among individual stable cell lines

*intellectual property of Originus, Inc. Ann Arbor, MI

For more information about STEP cell-based assays from Cayman, scan this code with your mobile device, call **(800) 364-9897** or visit:
www.caymanchem.com/STEP

Cayman CHEMICAL

RESEARCH

NEW ONLINE

Papers published this week at Nature.com/nature

NEWS & VIEWS Second generation galactic halo stars provide a glimpse of an earlier age **p.288**

ARTICLE 'Passenger' genes encode a range of new anticancer targets **p.337**

LETTER Native Americans derive from three waves of migration from Asia **p.370**

PUBLISHED 12 AUGUST 2012

CANCER

Drug targets in Burkitt lymphoma

Intensive chemotherapy can cure Burkitt's lymphoma, but its toxicity makes it unsuitable for more vulnerable patients. Roland Schmitz *et al.* identify mutations in the transcription factor TCF3 or its negative regulator ID3 in many sporadic Burkitt lymphomas, and suggests several drug targets, including PI(3) kinase and its downstream pathways.

Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics
R Schmitz *et al.*
doi:10.1038/nature11378

GENETICS

HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle
M A Deardorff *et al.*
doi:10.1038/nature11316

CANCER

Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling
M Dühren-von Minden *et al.*
doi:10.1038/nature11309

IMMUNOLOGY

Phosphorylation of NLRC4 is critical for inflammasome activation
Y Qu *et al.*
doi:10.1038/nature11429

PUBLISHED 15 AUGUST 2012

CHEMISTRY

A new route to prostaglandins

Prostaglandins are hormone-like messengers that regulate a broad range of physiological activities. Their biological activities and their complex molecular architectures have made prostaglandins popular targets for synthetic organic chemists for more than 40 years. Graeme Coulthard *et al.* now provide a concise synthesis of the most complex prostaglandin, $PGF_{2\alpha}$. Access to this compound should not only render existing prostaglandin-based drugs more affordable, but also facilitate the rapid exploration of the chemical space around the ubiquitous five-membered ring motif.

Stereocontrolled organocatalytic synthesis of prostaglandin $PGF_{2\alpha}$ in seven steps
G Coulthard *et al.*
doi:10.1038/nature11411

ENVIRONMENTAL SCIENCE

Putting a value on ocean well-being

There is much concern about the state of the oceans, both in terms of their importance to Earth's ecosystems and in their ability to provide benefits to society. Global ocean policy therefore strives for 'healthy oceans', yet we have no tool to actually measure ocean health or its response to interventions. To fill this gap, Benjamin Halpern *et al.* have developed an ocean health index that assesses the many factors that contribute to the health and benefits of the oceans. They determined scores for all coastal nations, and the global score is 60 of 100. Both uninhabited and developed coastlines score well, but many African and Asian countries score poorly.

An index to assess the health and benefits of the global ocean
B S Halpern *et al.*
doi:10.1038/nature11397

PHYSICS

A noisy quantum signature

Cavity optomechanics — in which light exerts a force on massive objects in a cavity through radiation pressure — is a highly active research field. Until now, it has not been possible to detect quantum aspects of this force, because of masking by technical and thermal noise. Now, Daniel Brooks *et al.* report a cavity optomechanical system consisting of ultracold atoms in which collective atomic motion is dominantly driven by quantum radiation pressure fluctuation. The back action of this motion onto the cavity light field produces a quantum effect called ponderomotive squeezing. The results could lead to applications such as subquantum-limited measurement in optomechanical sensors, fundamental tests of quantum mechanics at macroscopic scales and mechanical linkage between quantum devices.

Non-classical light generated by quantum-noise-driven cavity optomechanics
D W C Brooks *et al.*
doi:10.1038/nature11325

DISEASE

Dissecting malaria pathogenesis

This genome-wide association study in 2,645 cases and 3,050 controls from Ghana in West Africa identifies genetic variants that might affect any one of many steps in malaria pathogenesis. Two new loci are described. The first is near the *ATP2B4* gene on chromosome 1, encoding the main calcium pump of erythrocytes — host cells to the pathogenic stage of malaria parasites. The second is on chromosome 16, near *MARVELD3*, which encodes a tight-junction protein expressed on endothelial cells.

Genome-wide association study indicates two novel resistance loci for severe malaria
C Timmann *et al.*
doi:10.1038/nature11334

CANCER

Recurrent R-spondin fusions in colon cancer
S Seshagiri *et al.*
doi:10.1038/nature11282

Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors
B-K Koo *et al.*
doi:10.1038/nature11308

IMMUNOLOGY

Caspase-11 increases susceptibility to *Salmonella* infection in the absence of caspase-1
P Broz *et al.*
doi:10.1038/nature11419

NATURE PODCAST

Alison Abbott tells us about her trip to Kyrgyzstan with a team of Italian geneticists mapping food preferences along the Silk Road. We also discuss the return of the maser, and synthesizing prostaglandins in seven easy steps.
nature.com/nature/podcast

CANCER

Exploiting collateral damage

Some mutations in tumour cells play no part in causing cancer, but they generate cellular weak spots that may allow tumour cells to be selectively killed by drugs. SEE ARTICLE P.337

BEN LEHNER & SOLIP PARK

The central challenge in anticancer therapy is to kill tumour cells without harming other cells of the body. A defining property of cancer cells is that they carry genetic changes that distinguish them from normal cells and that should, at least in theory, allow them to be specifically killed by drugs that target them. But in practice this represents a formidable problem. On page 337 of this issue, Muller *et al.*[1] outline a potential strategy that exploits 'passenger' mutations. These mutations can remove genetic redundancies in cancer cells, making them vulnerable and so facilitating their destruction.

Most cancer research focuses on the genetic mutations that cause cancer — 'driver' mutations that typically activate cancer-causing oncogenes or inactivate tumour-suppressor genes. Such mutations make logical drug targets, and, in recent decades, an explosion of knowledge about their mechanism of action has led to the development of several anticancer drugs. Examples include the successful inhibition of the proteins encoded by the oncogenes *BCR-ABL*, *BRAF* and *ERBB2* in patients with chronic myelogenous leukaemia, melanoma and breast cancer, respectively[2,3]. But in many other cancers it has proved difficult to target the driver mutations effectively, even though they have been known for a long time.

An alternative approach would be to target the vulnerabilities created by passenger mutations. Although the development of a tumour may depend on only a small number of driver mutations, tumour cells usually also contain other genetic changes that accumulate as the tumours develop, often through exposure to environmental mutagens such as tobacco smoke and ultraviolet light[4]. Solid tumours can contain tens of thousands of such mutations[4]. The key point is that these passenger mutations also distinguish cancer cells from normal cells. So if the presence of a passenger mutation means that a cell becomes sensitized to the effects of a particular drug, then this should allow the tumour cells to be specifically killed.

Muller *et al.* outline one way in which this might be achieved. Many essential processes in biology are performed by two or more genes encoding proteins that have very similar biochemical functions. This 'partial redundancy' arises through the duplication of genes. It is a widespread phenomenon, and is maintained in genomes over vast evolutionary timescales[5]. One of the effects of partial redundancy is that the loss of one gene sometimes has little effect unless a second, related gene is also inhibited. Muller and colleagues' idea is that, if a passenger mutation removes or inactivates a duplicated gene, the tumour cells will become exquisitely sensitive to inhibition of the second, partially redundant gene (Fig. 1).

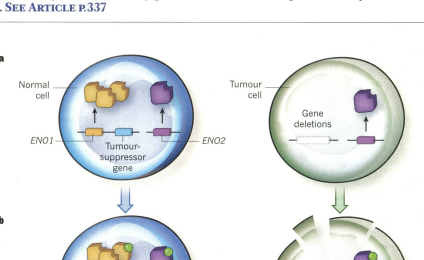


Figure 1 | Therapy in the passenger seat. The genomes of cancer cells often contain 'passenger' mutations, which are not directly related to the disease but which can create cellular vulnerabilities that allow the tumour cells to be killed by drugs targeting them. **a**, For example, Muller *et al.*[1] show that in the case of the brain cancer glioblastoma, a cancer-causing deletion of a tumour-suppressor gene also leads to loss of the *ENO1* gene, which encodes the enzyme enolase. This makes the tumour cells dependent on *ENO2*, which also encodes enolase but is expressed at lower levels than *ENO1* in brain cells. **b**, Normal cells will survive treatment with a drug that inhibits enolase activity (indicated by the green circle), because of their high expression of *ENO1*. The low levels of enolase in the tumour cells, however, mean that they will be killed by this treatment.

To test this idea, the authors hunted for passenger mutations that resulted in the deletion of a duplicated gene in a type of brain cancer called glioblastoma. They found that, in their tumour samples, the gene *ENO1* was deleted in several cases — most probably because it is located in a region of the genome that also contains one or more tumour-suppressor genes that, when deleted, promote uncontrolled proliferation of the tumour cells. *ENO1* is one of three human genes that encode enolase, an enzyme required for the metabolic pathway that converts sugar to energy. These genes originated from duplications of an ancestral enolase gene more than 500 million years ago, and each gene is expressed to varying extents in different tissues. In the brain, for example, most enolase activity is provided by the ENO1 protein, with a smaller contribution from ENO2.

Consistent with their hypothesis, the authors found that glioblastoma cells in which *ENO1* is deleted are much more sensitive to inhibition of *ENO2* than other cells. They tested this sensitivity in cultured cells using two methods of inhibition — a known small-molecule inhibitor of enolase and a short hairpin RNA

molecule (shRNA) that binds specifically to enolase-encoding messenger RNAs and interferes with the production of enolase proteins. The authors found that in normal cells, because of the high levels of ENO1 protein present, *ENO2*-specific shRNA molecules or the enolase inhibitor had little effect. But in glioblastoma cells in which *ENO1* was deleted, both treatments killed the cells (Fig. 1). Indeed, when the authors inhibited *ENO2* activity in the *ENO1*-deficient glioblastoma cells, the cells could no longer form tumours when transferred into mice.

The study therefore provides a powerful proof of concept that the loss of redundancy in cancer genomes can be exploited to selectively kill tumour cells. In glioblastoma, the strategy works because the deleted duplicate gene (*ENO1*) normally provides most of the enolase activity, and there seems to be no compensatory increase in protein production from the second gene (*ENO2*). But compensatory changes in gene expression from duplicate genes may actually be common[6,7], and in other types of cancer the overall difference in protein activity between the tumour and normal cells may be much harder to target therapeutically. Moreover, the strategy must still be tested in animal models and in patients, particularly as a combination therapy with other drugs, because compensatory increases in the expression of the remaining gene duplicate may lead to the rapid development of drug resistance.

Finally, although not mentioned by Muller and colleagues, the vulnerabilities created by common passenger mutations are likely to extend far beyond the loss of protein redundancy[8]. Systematic screens to identify 'synthetic lethal' interactions[9,10], in which a drug or a gene inhibition kills only cells carrying a common passenger mutation, may be one way to identify these additional vulnerabilities. For instance, another recent study[11] presents an alternative strategy for targeting passenger mutations. It takes advantage of situations in which the activity of a single gene is reduced in cancer cells, for example through deletion of one of the two copies of the gene, rendering the cells highly sensitive to further inhibition of the same gene.

Muller and colleagues' results provide an impetus for two major conceptual advances: first, that the loss of protein redundancy provides a therapeutic opportunity to kill specific cells, and second, that passenger mutations may be the Achilles heel of cancer genomes. In our opinion, this second idea is the more important, and the plethora of passenger mutations present in cancer genomes should create many opportunities for personalized therapies. This underlines how crucial it is that, rather than simply focusing on cancer-causing genes and pathways, researchers also consider the therapeutic opportunities created by common passenger mutations. ■

Ben Lehner *and* **Solip Park** *are in the EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation, Universitat Pompeu Fabra, 08003 Barcelona, Spain.* **B. L.** *is also at the Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona.*
e-mail: ben.lehner@crg.eu

1. Muller, F. L. *et al. Nature* **488,** 337–342 (2012).
2. Sellers, W. R. *Cell* **147,** 26–31 (2011).
3. Workman, P. & de Bono, J. *Curr. Opin. Pharmacol.* **8,** 359–362 (2008).
4. Stratton, M. R., Campbell, P. J. & Futreal, P. A. *Nature* **458,** 719–724 (2009).
5. Vavouri, T., Semple, J. I. & Lehner, B. *Trends Genet.* **24,** 485–488 (2008).
6. Kafri, R., Bar-Even, A. & Pilpel, Y. *Nature Genet.* **37,** 295–299 (2005).
7. Burga, A., Casanueva, M. O. & Lehner, B. *Nature* **480,** 250–253 (2011).
8. Lehner, B. *Trends Genet.* **27,** 323–331 (2011).
9. Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Friend, S. H. *Science* **278,** 1064–1068 (1997).
10. Chan, D. A. & Giaccia, A. J. *Nature Rev. Drug Discov.* **10,** 351–364 (2011).
11. Nijhawan, D. *et al. Cell* http://dx.doi.org/10.1016/j.cell.2012.07.023 (2012).

APPLIED PHYSICS

Masers made easy

The technological potential of masers — the microwave equivalents of lasers — has been thwarted by their impractical operating conditions. A solid-state maser that works at room temperature may change that. SEE LETTER P.353

AHARON BLANK

Noise is an integral part of any electronic system, and it comes in many forms. We are all familiar with the noise that emerges from our radio when we try to listen to a distant station, or the 'snow-like' noise that appears on a detuned television screen. Scientists and engineers have been battling with noise for years, trying to make radio, television and mobile-phone communication better and clearer. In terms of noise reduction, however, the best performance is achieved only when electronic systems are operated at cryogenic temperatures. On page 353 of this issue, Oxborrow *et al.*[1] describe a solid-state microwave source and amplifier that has an exceptionally low noise level and that operates at room temperature. The device may find applications in space communication, radio astronomy and microwave spectroscopy.

Sources of noise can be roughly divided into those that are an intrinsic part of the signal of interest, about which not much can be done, and those that are added to the signal by external components, such as systems of electronic detection, reception or amplification. These extrinsic sources can be minimized, but the proper handling of signals and noise depends greatly on their characteristic frequency. One of the most crucial but challenging frequency bands for signal detection and amplification is the micro- and millimetre-wave range — frequencies of 1–100 gigahertz (1 gigahertz is 10^9 Hz). This part of the electromagnetic spectrum is commonly used for: surveillance radars; communication for mobile phones; space communication; radio astronomy (including the search for signals coming from extraterrestrial intelligence); and a variety of microwave-spectroscopy techniques.

In the first two of these applications, the noise levels that come with the signal are quite high. This means that low-cost amplifiers and detectors can be used at room temperature without significantly increasing the high intrinsic noise levels, and thus with almost no deterioration in the quality of the measured signals. However, in the other applications mentioned, the intrinsic noise can be very low. This means that considerable sensitivity to weak signals would be gained if the electronic detection and amplification of these signals were almost noiseless.

One of the earliest methods for amplifying faint signals in the microwave range, and still the best one, is based on the use of a maser — a device that performs 'microwave amplification by stimulated emission of radiation'. Masers are the predecessors and radio equivalents of lasers. They were at the centre of at least two Nobel prizes, one awarded to Charles Townes in 1964 for inventing the device, and the other to Arno Penzias and Robert Wilson in 1978 for the discovery of cosmic microwave background radiation — relic radiation from the Big Bang. The latter achievement was made possible only by the use of a maser amplifier whose noise contribution to the signal the researchers were detecting was so small that they had no choice but to 'blame' the background radiation for the noise in their system[2].

Masers of the type used in these experiments are based on crystals 'doped' with a paramagnetic substance — one that contains unpaired electron spins. In these systems, most of the unpaired electrons are found at the lowest energy level of the medium. But if the crystal is 'pumped' with microwave radiation, it is possible to reach a situation in which there are more of these electrons in an excited state than in a lower-energy state (a mechanism known

as population inversion) and to 'stimulate' them to decay into this lower state and emit radiation of the same frequency and phase as the incident radiation (Fig. 1a). In this way, the microwave radiation passing through the crystal is amplified. This emission and amplification is analogous to that of lasers, but occurs at much lower frequencies.

In such masers, the intrinsic noise, which is mainly caused by spontaneous photon emission, is very low — close to its physical limit. So similar designs are still used today in fundamental science, as well as in advanced space communications and radio-astronomy applications. However, for population inversion to be attained, the masers need to be cooled to liquid-helium temperatures (4.2 kelvin). This, together with their limited output power and the fact that their operation requires a rather large static magnetic field, has led to masers becoming a rare, almost extinct, species. Furthermore, in recent years, microwave technology has advanced so much that the noise levels achieved by cryogenically cooled, conventional semiconductor-based or superconducting amplifiers is comparable to that of masers, but with a much better power and bandwidth performance and far less physical complexity[3,4].

Now Oxborrow and colleagues have solved one of the most troubling aspects of solid-state maser operation. They demonstrate a device working at room temperature — a feat not previously achieved even though it has been five decades since the invention of the solid-state maser. They did this by using a special system: a crystal of the organic compound *p*-terphenyl doped with pentacene molecules. The energy levels of a pentacene molecule can be pumped optically to achieve large population inversion even at room temperature (Fig. 1b). A large population inversion, which leaves few electrons at the lower energy level, is required for the amplification process, and reduces the intrinsic noise of the maser.

In addition, Oxborrow *et al.* operated the maser using a microwave resonator that has low energy loss. Such low-loss resonators ensure that the device functions as an amplifier rather than as a sophisticated, but expensive, attenuator. In principle, their maser can be used to amplify weak microwave signals that have small intrinsic noise and without the need for any cooling. On the downside, like the 'old-fashioned' solid-state masers, its frequency of operation can be tuned only by an external magnetic field. However, this operational aspect is far less complicated than cryogenic cooling. Several other microwave devices, such as YIG oscillators, which also use external magnetic fields for frequency tuning, are in wide commercial use.

Finally, the authors' optically pumped maser delivers pulses of microwave radiation, rather than continuous signals. Ideally, amplifiers and sources of radiation should be able to operate in a continuous mode, which is more useful than 'part-time' operation. Nevertheless, with additional improvements — such as size reduction, a more efficient light-excitation protocol and the use of several resonators in parallel — the device may be useful for the applications mentioned previously. For example, the new maser can be seen as a first step towards a space-communication capability with which we could talk to extraterrestrials using 'phones' at room temperature. ■

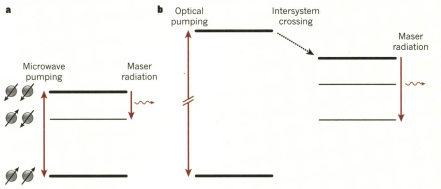


Figure 1 | Pumping masers. a, A conventional solid-state maser has at least three energy levels (horizontal lines), corresponding to the three possible spin combinations of two unpaired electrons (spheres; black arrows indicate spin states). Pumping the system's atomic medium with microwave radiation brings the electron spins from the lowest energy level to an excited state until the electron population of the levels is equalized. Subsequent decay of the electrons to the intermediate state results in the emission of maser radiation. The thickness of the lines denotes each level's population. **b**, In Oxborrow and colleagues' maser[1], electrons in pentacene molecules within a host crystal are optically pumped and undergo 'intersystem crossing', whereby they decay to a triplet state in which the uppermost level is the most populated. Decay from this level to the triplet's lowest level is accompanied by the emission of maser radiation.

Aharon Blank *is in the Schulich Faculty of Chemistry, Technion-Israel Institute of Technology, Haifa 32000, Israel.*
e-mail: ab359@tx.technion.ac.il

1. Oxborrow, M., Breeze, J. D. & Alford, N. M. *Nature* **488,** 353–356 (2012).
2. Penzias, A. A. & Wilson, R. W. *Astrophys. J.* **142,** 419–421 (1965).
3. Schleeh, J. *IEEE Electron. Device Lett.* **33,** 664–666 (2012).
4. Eom, B. H., Day, P. K., LeDuc, H. G. & Zmuidzinas, J. *Nature Phys.* **8,** 623–627 (2012).

CELL BIOLOGY

Neither brown nor white

Fat cells are usually thought of as being either energy-storing white fat cells or food-burning brown fat cells. The identification of a third type of fat cell in mice and humans might open up new avenues for combating obesity.

BARBARA CANNON & JAN NEDERGAARD

In humans as in other mammals, fat is stored in white fat cells, whereas brown fat cells burn fat to generate heat. However, it seems that not everything is white or brown. Writing in *Cell*, Wu *et al.*[1] present convincing evidence that a distinct brown-like type of adipocyte (fat cell) exists within white fat depots.

Research on brown adipose tissue is developing rapidly. Until recently, it was thought that brown fat cells were merely modified white adipocytes. We now accept that the heat-producing, and thus food-consuming, brown fat cells are in reality muscle relatives[2–4] (Fig. 1a). Both muscle cells and brown fat cells have many mitochondria — organelles often referred to as the cell's powerhouses — that burn food to liberate large amounts of energy. Muscle cells use this energy for the generation of ATP molecules that power muscle contraction. By contrast, brown adipocytes release the energy directly as heat, a uniquely mammalian ability gained through the development of the mitochondrial protein UCP1 (uncoupling protein 1). A deficiency of UCP1 leads not only to a lack of heat production but also to an increased propensity for

obesity, at least in mice[5]. And, importantly, although it was thought that in humans active brown fat existed only in babies and was progressively lost later in life, it is now clear that adults have active brown adipose tissue[6–11], and that the absence of this tissue might even be functionally linked to obesity[8].

When animals are chronically exposed to cold, the ability of brown fat to produce heat is enhanced manyfold, owing to an increase in the total amount of UCP1 in the tissue. In addition, certain apparently white adipose depots can undergo a shift in colour from whitish to brownish[12–14]. This 'browning' process is accompanied by a large increase in the expression of the UCP1 gene, relative to the very low starting level. A similar browning can be promoted in animals by various types of genetic manipulation. Earlier, when it was thought that brown fat cells were just some kind of modified white fat cell, the browning phenomenon seemed understandable. However, because we now know that brown fat cells are related to muscle and not to white fat cells, it is unlikely that the UCP1-expressing cells that appear in white fat after browning are classical brown fat cells. Indeed, previous studies[13–15] have suggested the existence of a subpopulation of cells in white fat depots that have the ability to express UCP1, and therefore to become brownish.

To explore whether such a subpopulation indeed exists, Wu *et al.* isolated progenitor cells (cells that can give rise to specialized cell types) from the subcutaneous white adipose tissue of mice and 'immortalized' these cells. Through this procedure, the cells acquired the ability to proliferate indefinitely and so yielded stable cell lines. Among hundreds of such cell lines, about 30 could be induced to accumulate fat.

The authors then analysed the gene-expression profiles of these cell lines. About a third of the cell lines showed an expression profile that was similar, although not identical, to that of brown adipocytes, whereas the rest displayed a quite different but rather homogeneous gene-expression pattern. Therefore, the original, apparently white, adipose tissue depot contained progenitors of genuine white fat cells, as well as a subpopulation of progenitors of a different type of adipocyte that — given an adequate stimulus — could express UCP1 and thus generate heat: the 'beige' or 'brite' (brown-in-white) cells (Fig. 1b). The nomenclature for these fascinating cells is presently unsettled, and they are also known as 'ectopic brown fat cells' or 'inducible brown fat cells'.

These results raise several questions. What is the origin of the two types of cell? Are there two distinct types of pericytes (cells that are wrapped along the blood capillaries and supposedly give rise to fat cells)? Or does one cell type derive, for example, from the bone marrow? How can certain white-like adipocytes, which in general possess very few mitochondria, suddenly enhance their mitochondrial complement during the 'beiging/britening' process?

Moreover, there is an additional twist to Wu and colleagues' story — and it concerns humans. The authors identified a group of mouse genes that were characteristically expressed in either beige/brite or brown fat cells. When they examined the gene-expression profile of brown fat from 11 adult humans, they unexpectedly found that the profile was closer to that of mouse beige/brite cells than to that of mouse brown cells. So, the brown fat of adult humans might not really be 'brown', but might instead be equivalent to mouse beige/brite adipocytes.

This suggestion is surprising in several ways. Brown fat in adult humans[6–11] is found in practically the same regions of the body as it occurs in babies. As described by Aherne and Hull in a classic paper[16] in 1966, brown adipose tissue in babies is present partly as a thin interscapular sheet between the shoulder blades (this sheet, as we now know, is not detectable in adults) but mainly as many smaller masses around the muscles and blood vessels of the neck — extensions of these masses pass under the clavicles to form rather large deposits in the axillae. Do Wu and colleagues' observations nevertheless mean that humans, even when we are babies, do not have genuine brown fat (that is, of the same type as that of other mammals)? Or is brown fat exchanged for beige/brite fat at some time during early human life? And is there a principal difference between the white (subcutaneous) adipose tissue in mice and humans? How should we modify our current mouse models to make them more similar to humans with respect to adipose tissues? And finally, can the beiging/britening process be stimulated so that we can obtain more food-burning cells that could help to combat obesity? ■

Barbara Cannon *and* **Jan Nedergaard** *are at The Wenner-Gren Institute, Stockholm University, Stockholm SE-106 91, Sweden.*
e-mail: barbara.cannon@wgi.su.se

1. Wu, J. *et al. Cell* **150**, 366–376 (2012).
2. Timmons, J. A. *et al. Proc. Natl Acad. Sci. USA* **104**, 4401–4406 (2007).
3. Seale, P. *et al. Nature* **454,** 961–967 (2008).
4. Cannon, B. & Nedergaard, J. *Nature* **454,** 947–948 (2008).
5. Feldmann, H. M., Golozoubova, V., Cannon, B. & Nedergaard, J. *Cell Metab.* **9**, 203–209 (2009).
6. Nedergaard, J., Bengtsson, T. & Cannon, B. *Am. J. Physiol.* **293**, E444–E452 (2007).
7. Cypess, A. M. *et al. N. Engl. J. Med.* **360**, 1509–1517 (2009).
8. Saito, M. *et al. Diabetes* **58,** 1526–1531 (2009).
9. van Marken Lichtenbelt, W. D. *et al. N. Engl. J. Med.* **360**, 1500–1508 (2009).
10. Virtanen, K. A. *et al. N. Engl. J. Med.* **360,** 1518–1525 (2009).
11. Zingaretti, M. C. *et al. FASEB J.* **23**, 3113–3120 (2009).
12. Young, P., Arch, J. & Ashwell, M. *FEBS Lett.* **167,** 10–14 (1984).
13. Cinti, S. *et al. J. Histochem. Cytochem.* **50,** 21–31 (2002).
14. Guerra, C. *et al. J. Clin. Invest.* **102,** 412–420 (1998).
15. Petrovic, N. *et al. J. Biol. Chem.* **285,** 7153–7164 (2010).
16. Aherne, W. & Hull, D. *J. Pathol. Bacteriol.* **91,** 223–234 (1966).

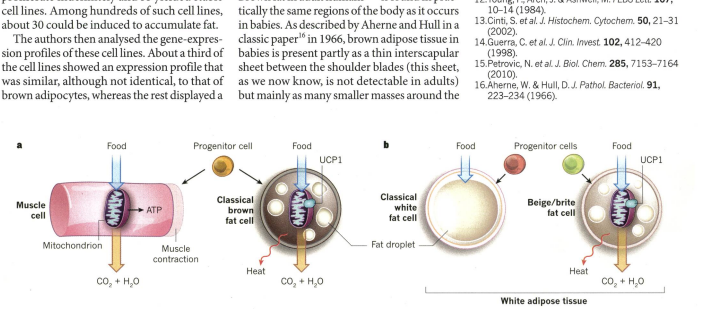


Figure 1 | Different progenitors, different colours. a, Cells that make up muscle and brown fat are derived from the same progenitor cells. Muscle cells and classical brown fat cells contain many mitochondria, organelles in which food is burned to produce carbon dioxide and water. The energy released is harvested in the form of ATP molecules that support contraction in muscles, whereas it is directly released as heat in brown fat. This heat generation requires the presence of the UCP1 protein. **b,** Wu *et al.*[1] identified two types of progenitor cell in mouse white adipose tissue. These cells give rise to classical white fat cells (which store energy in the form of fat droplets) and to 'beige' or 'brite' fat cells. Beige/brite cells express UCP1 and are functionally similar to brown cells, in spite of being derived from different progenitors.

ASTROPHYSICS

One of the first of the second stars

The chemical content of a star that was born relatively shortly after the formation of the Milky Way calls into question conventional understanding of how stars formed in the early Universe.

JOHN COWAN

How did the first stars form early in the history of our Galaxy and the Universe, and what were these stars like? The Big Bang produced only hydrogen, helium and some lithium, so the first stars would have contained only those elements. Given that the stars became extinct long ago, they were probably quite massive — the more massive a star (with masses more than 80 times that of the Sun a possibility), the shorter its lifetime (as little as a few million years). However, we can learn much about the early conditions in our Galaxy, including the types of element formed and the nature of these first stars, by studying surviving stars from a second generation of stars, which formed from the debris of the first. These surviving 'second stars', which are less massive than the Sun and live for many billions of years, can be found in the Galactic halo. Writing in *Astronomy & Astrophysics*, Caffau *et al.*[1] describe a Galactic halo star that is one of the earliest members of this second generation.

This new work is an extension, and expansion, of Caffau and colleagues' initial report[2] on this star, which is catalogued in the Sloan Digital Sky Survey and known by the nonsexy name of SDSS J102915+172927. The authors used the Very Large Telescope (VLT) in Chile to undertake a detailed spectroscopic analysis[1] of the elements contained in the star, which is in the centre of the constellation Leo. They report that the star has an extremely low abundance of iron — approximately 1/130,000 that of the Sun. Astronomers refer to a star's iron abundance as its metallicity. This metallicity generally correlates with Galactic time: iron is produced in exploding stars, known as supernovae, and its abundance in our Galaxy has increased progressively over time. Thus, the most iron-deficient — or extremely metal-poor (EMP) — stars are among the oldest, having formed relatively soon after the Galaxy itself, early in the history of the Universe.

And yet SDSS J102915+172927 does not have the lowest known iron abundance. Two other stars[3,4] are even more iron-poor, implying that they are older than the star in Leo. Compared to the Sun, however, both stars have a large abundance of carbon and nitrogen relative to iron, as has been noted in several other EMP stars. By contrast, Caffau *et al.*[1] found that carbon and nitrogen were not enhanced in SDSS J102915+172927 — the abundance of these elements with respect to that of iron is consistent with solar values. Similarly to some other EMP stars[4], the abundance of oxygen could not be measured in SDSS J102915+172927. However, Caffau and colleagues also defined a metal-mass fraction based on the total abundance of all of the elements heavier than helium, rather than just that of iron. They report that SDSS J102915+172927 has the lowest such value ever measured, and argue that this makes its composition similar to that of the primordial gas that existed shortly after the Big Bang.

The types and abundances of these elements in old, low-metallicity stars are crucial to our understanding of what happened before the stars' formation. First, it is a puzzle how a low-mass star such as SDSS J102915+172927 (which is less massive than the Sun) even formed early in the history of the Galaxy at a time when high-mass stars would seem to be more common. Observational and theoretical studies[5,6] have suggested that elements such as carbon or oxygen are necessary for cooling (low-mass) parent gas clouds sufficiently for them to eventually collapse and form low-mass stars. The difference between SDSS J102915+172927, which has a relatively low carbon abundance, and carbon-enhanced metal-poor (CEMP) stars calls into question what is normal for these early stars[7]. But with so few stars of this type observed, it is difficult to discern a pattern.

Perhaps SDSS J102915+172927 is older than the other EMP stars observed: the relationship between iron abundance and time might not be entirely linear so early on, and thus the lowest metallicity star might not be the oldest. SDSS J102915+172927 might fall into the transition region between the first generation of stars (sometimes referred to as Population III) and the second generation, or Population II; halo, EMP and CEMP stars belong to the latter group.

Or perhaps SDSS J102915+172927 formed in a region of the Galaxy that had particularly low levels of elements heavier than helium. Although carbon can be produced internally in ageing (giant) stars, this star is probably an (unevolved) main-sequence star — a stellar phase similar to the current state of the Sun — and not a giant or a sub-giant[2]. This means that the carbon and other heavy elements observed in SDSS J102915+172927 must have been synthesized in a supernova and then incorporated into the gas that would form new stars.

The chemical-abundance pattern observed in SDSS J102915+172927 is consistent with what is predicted for such a supernova event[2]. Yet this star in Leo does not have enhanced carbon, in sharp contrast to one of the unevolved carbon-enhanced EMP stars, which also has detectable strontium[8] — a rare heavy element made only in supernovae by a series of neutron captures.

A large scatter in the abundance of heavy, neutron-capture elements with respect to iron is observed for metal-poor stars. (There are many metal-poor halo stars but only a few EMP stars.) This scatter suggests that, at early times, the Galaxy was an unmixed, or non-homogeneous entity, with individual element-synthesis events — that is, supernovae — scattered throughout the halo[8]. Thus, the greatly varying carbon abundances in these early stars might also result from this heterogeneity. Clearly, observations of additional stars will be needed to probe this early phase of the Milky Way.

Further support for this early history of the Galaxy was provided by Caffau and colleagues' measurement[1,2] of the lithium abundance in SDSS J102915+172927. Lithium, produced in the Big Bang, is expected to be at a uniform and primordial abundance level (denoted as the Spite plateau[9]) in these early stars. Surprisingly, however, the measured lithium abundance in this, and another[4], EMP star is low, below the observed Spite plateau. This suggests, at least for some of these early stars, that there is probably a lithium-destruction mechanism occurring during star formation. It will be crucial to find additional evidence of variations in lithium abundance to better understand the earliest stages of star formation.

There will be more stars such as SDSS J102915+172927. Only a small fraction of the thousands of stars in the SDSS database has been observed with the VLT. Nevertheless, this new discovery is a valuable first step in filling in the gaps in our knowledge of the early history of the Universe, and of how stars and elements were formed. ■

John Cowan *is in the Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73019, USA.*
e-mail: cowan@nhn.ou.edu

1. Caffau, E. *et al. Astron. Astrophys.* **542,** A51 (2012).
2. Caffau, E. *et al. Nature* **477,** 67–69 (2011).

3. Christlieb, N. *et al. Nature* **419,** 904–906 (2002).
4. Frebel, A. *et al. Nature* **434,** 871–873 (2005).
5. Bromm, V. & Larson, R. B. *Annu. Rev. Astron. Astrophys.* **42,** 79–118 (2004).
6. Frebel, A., Johnson, J. L. & Bromm, V. *Mon. Not. R. Astron. Soc.* **380,** L40–L44 (2007).
7. Norris, J. E. *Nature* **477,** 37–39 (2011).
8. Sneden, C., Cowan, J. J. & Gallino, R. *Annu. Rev. Astron. Astrophys.* **46,** 241–288 (2008).
9. Spite, M. & Spite, F. *Nature* **297,** 483–485 (1982).

NEUROSCIENCE

Circuits drive cell diversity

Neurons of the same type can show functional differences. It turns out that this diversity is in part the result of the cells' adaptation to their specific neural networks. SEE LETTER P.375

NATHANIEL URBAN & SHREEJOY TRIPATHY

Modern manufacturing was revolutionized by the use of interchangeable parts so similar in their function that any one could effectively replace any other. Making such parts meant that manufacturers did not need to keep track of which nut worked with which bolt or which piston was intended to go in which cylinder. Similarly, most physiological analyses ignore cell-to-cell variation and focus instead on differences between cell types, as though each cell of a specific class were functionally equivalent to any other cell of the same type. However, neuroanatomists have long marvelled at the snowflake-like diversity apparent in the shapes of individual neurons, even within a cell type. And recent analyses have demonstrated that same-class neurons show substantial heterogeneity in their intrinsic properties[1–3], although the origin of such diversity is poorly understood. On page 375 of this issue, Angelo *et al.*[4] report that physiological variability among mitral cells (a type of neuron in the olfactory system) is at least partly caused by differences in the inputs that they receive.

Sensory neurons in the nose are activated when odorant molecules bind to specific receptor proteins on the neurons' surface. This activation is transmitted as an excitatory signal along nerve fibres (axons) that terminate in structures called glomeruli, which cover the surface of the brain's olfactory bulb. Each of about 2,000 glomeruli in the mouse olfactory bulb receives axons from sensory neurons that express a single type of odorant receptor (Fig. 1).

Mitral cells receive these excitatory signals from the sensory neurons and are the main source of the olfactory bulb's output to the cerebral cortex; of note, each mitral cell is connected to a single glomerulus. Although they have long been considered to be a single class of neuron, mitral cells show substantial variation in their physiological properties[1,2]. This is probably because of differences in the expression of certain ion channels — pores in the cell membrane that mediate the transmission of electrical signals by allowing the passage of specific ions. Indeed, Angelo *et al.* have previously observed[1] that mitral cells display considerable diversity in the magnitude of a transmembrane current known as a 'sag'. This current, which is generated when the resting electric potential across the neuron's membrane is intensified, is believed to regulate the sensitivity and timing of mitral cells' responses to periodic sensory inputs (such as those generated when an animal sniffs rapidly).

To explore the causes of variability in sag currents, the authors simultaneously recorded the sag from pairs of mitral cells belonging to either the same or different glomeruli, in slices of mouse olfactory bulb. Although sag magnitude was highly variable from one cell to the next, the variability was markedly lower between mitral cells that received input from the same glomerulus than between cells that received input from different glomeruli.

Why are same-glomerulus mitral cells so homogeneous? The hypothesis explored by Angelo *et al.* is that neurons with similar levels of activity acquire similar physiological features. Pairs of mitral cells connected to the same glomerulus have the same sources of input (Fig. 1) and will therefore have more similar activity levels than random pairs of mitral cells. If the expression of sag-mediating ion channels is affected by a neuron's activity, then this could account for the relative homogeneity of mitral cells connected to the same glomerulus.

To address this issue, the authors cleverly

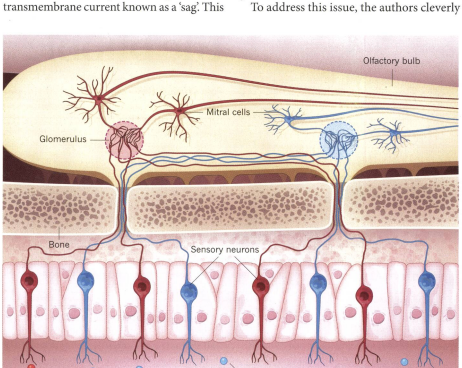


Figure 1 | The importance of network affiliation. Sensory neurons in the nose are activated by specific odorants or groups of odorants (different types are indicated by different colours), and transmit this information to mitral cells, another type of neuron in the brain's olfactory bulb. The transmission takes place in spherical structures called glomeruli, each of which receives input from a single type of sensory cell. Angelo *et al.*[4] report that mitral cells that are linked to the same glomeruli are functionally more similar to each other than to mitral cells that connect to other glomerular networks. This result suggests that input heterogeneity drives functional diversity of mitral cells.

50 YEARS AGO

A REPORT entitled *The Supply of Science Teachers in England and Wales* ... states that, immediately, there is need for an increase of between 28,000 and 34,000 over the present number of graduate science teachers in schools, technical colleges, training colleges and universities, which stands to-day at about 20,600 ... The shortage of science teachers is seen to be one part of the general teacher shortage, but many reasons are given for regarding it as a separate and specially urgent problem.
From *Nature* 18 August 1962

100 YEARS AGO

In 1904 Dr. J. R. Ashworth and I published ... observations on aged individuals of *Sagartia troglodytes* then and still in the possession of Miss Jessie Nelson in Edinburgh. After eight years these anemones are still in excellent health, having been in captivity for considerably more than half a century. In one respect I fear we did them an injustice, namely in attributing cannibalism to them, the error being probably due to the observation of the birth of young from a parent the tentacles of which were not fully expanded. Recently I chanced to notice a young Sagartia attached to a small piece of seaweed floating free in the aquarium. A slight agitation of the water was sufficient to bring the young anemone in contact with the tentacles of one of the patriarchs of its own species. They immediately closed round it and a small part of the disk became emarginate. The greater part, however, was not sensibly affected, and the mouth remained closed. In less than two minutes the folded-in tentacles uncurled and the young anemone was thrust away with some force ... Neither the young one nor the tentacles that held it were apparently affected in any way.
From *Nature* 15 August 1912

made use of an existing transgenic mouse strain, the 'monoclonal nose' mouse, in which almost all olfactory sensory neurons express the same odorant receptor. In these animals, most mitral cells receive similar inputs (regardless of their glomerular association) and would therefore be expected to show similar sag levels. Indeed, Angelo *et al.* found that randomly selected pairs of mitral cells in these mice showed much less diversity in sag magnitude than did equivalent pairs in control animals, consistent with the hypothesis that sag variation reflects heterogeneity in cellular activity.

One crucial question that the authors did not address is: what is the functional role, if any, of neuronal diversity? Theoretical studies have demonstrated that diversity could enhance information coding[2,5] and neural synchronization[6]; such studies provide valuable guidance for understanding how cellular diversity might support, or impair, brain function. This may tempt us to think that the robustness and power of brains as computational devices derive in part from the bespoke nature of individual neurons. However, a definitive answer to the question of whether neural variability is functionally important, or simply an unavoidable consequence of the imprecision of biological systems, awaits future experiments in which diversity can be controlled while probing behaviour. ■

Nathaniel Urban *and* **Shreejoy Tripathy**
are in the Department of Biological Sciences and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.
e-mail: nurban@cmu.edu

1. Angelo, K. & Margrie, T. W. *Sci. Rep.* **1,** 50 (2011).
2. Padmanabhan, K. & Urban, N. N. *Nature Neurosci.* **13,** 1276–1282 (2010).
3. Marsat, G. & Maler, L. *J. Neurophysiol.* **104,** 2543–2555 (2010).
4. Angelo, K. *et al. Nature* **488,** 375–378 (2012).
5. Stocks, N. G. *Phys. Rev. Lett.* **84,** 2310–2313 (2000).
6. Longtin, A. & Mejias, J. F. *Phys. Rev. Lett.* **108,** 228102 (2012).

OCEANOGRAPHY

The trouble with the bubble

Past studies have suggested that the ocean's nitrogen budget has a deficit of fixed nitrogen. This view may now change, thanks to upward revisions of the rate of nitrogen input through biological activity. SEE LETTER P.361

ANGELICQUE E. WHITE

As is the case for gardens, forests and fields, the availability of fixed nitrogen (such as nitrate and ammonia) can limit the productivity of our seas. Oceanographers are therefore interested in the relative magnitudes of sources and sinks of fixed nitrogen as dynamic controls of ocean fertility. The trajectory of the oceanic nitrogen inventory has been a long-debated and contentious topic: direct rate measurements of biologically mediated nitrogen fluxes suggest that the ocean is currently being depleted of this resource, whereas geochemical evidence indicates a steady-state, balanced nitrogen budget[1]. On page 361 of this issue, Großkopf *et al.*[2] report that applying a new twist to an old method may reconcile these views and lead to a revision of our understanding of the present-day oceanic fixed-nitrogen budget*.

Unlike other elemental cycles, which are predominantly influenced by riverine, atmospheric or sedimentary fluxes (such as the iron and phosphorus cycles), the oceanic inventory of fixed nitrogen is largely set by the push and pull of two biological processes. First, fixed nitrogen is added to the oceans by microorganisms known as diazotrophs, which convert the nearly limitless supply of atmospheric nitrogen gas (N_2) dissolved in sea water to ammonia. Second, in sediments and in oxygen-depleted zones of the ocean, fixed nitrogen is chemically reduced to N_2 by the microbial processes of denitrification and anaerobic ammonia oxidation. Understanding the relative balance of the fluxes of fixed nitrogen in the ocean requires the rate measurements of these competing processes to be accurate and well constrained.

On the source side of this nitrogen budget, the majority of published estimates of marine N_2-fixation rates[3] are based on a fairly straightforward protocol[4]: add a bubble of isotopically labelled nitrogen gas ($^{15}N_2$) to a sample of sea water; calculate the initial enrichment of $^{15}N_2$ in the sample using the ideal-gas law; incubate the sample for a specified period of time; and then measure the

*This article and the paper under discussion[2] were published online on 8 August 2012.

fraction of the isotopic tracer that is incorporated into cellular material in the sample. The net rate of N_2 fixation is then calculated by multiplying the fraction of biomass labelled per unit of time by the mass of nitrogen in particles suspended in the sample (which are assumed to be living material).

When these rates are extrapolated from localized study regions to the expanse of ocean thought to be habitable to diazotrophs, the resulting global N_2-fixation rates are generally found to be far less than those inferred by geochemical approaches[1]. This potential discrepancy set scientists on a "relentless search"[5] for potentially missing sources of fixed nitrogen. These efforts paid off — they led to the discovery of a wider cast of nitrogen-fixing 'characters' than had been known before[6]; an expanded view of the ocean habitats occupied by such organisms[7]; the elucidation of previously unknown metabolic strategies for diazotrophic growth[8]; and descriptions of newly recognized symbioses between diazotrophs and other microorganisms[9]. Despite these revelations, the available N_2-fixation-rate measurements still indicated a deficit of oceanic fixed nitrogen. And so the question remained: if the nitrogen budget is in balance, then where have rate measurements of oceanic nitrogen fluxes gone wrong?

In 2010, a potential solution to this conundrum was reported[10] with the demonstration that when $^{15}N_2$ is added to a sample of sea water, it does not rapidly dissolve in the surrounding liquid. This was an unexpectedly simple solution — if a fraction of the added $^{15}N_2$ remains in the gas phase during fixation-rate measurements, instead of dissolving in sea water as was assumed, then one would underestimate N_2-fixation rates. To assess the effect of incomplete $^{15}N_2$ dissolution, laboratory experiments were performed in which the classic method was revised[10]: instead of adding a bubble of $^{15}N_2$ directly to a sample, the tracer was added to aliquots of filtered and degassed sea water, which were agitated until the bubble dissolved fully. The isotopic enrichment of the water was then verified by mass spectrometry[10]. By adding a known volume of this water to samples, rather than a bubble of $^{15}N_2$, the initial concentration of added tracer could be constrained, leading to more-accurate measurements. This methodological tweak resulted in significantly higher fixation rates being measured in the laboratory — a finding that, if applicable to *in situ* measurements, could substantially reduce the apparent imbalances previously observed for the marine fixed-nitrogen budget.

Großkopf and colleagues now report the first application of the modified 'dissolution' method to the analysis of natural populations of diazotrophs. During two research cruises in the Atlantic Ocean, the authors made 17 side-by-side comparisons of N_2-fixation rates obtained using the conventional bubble-addition approach[4] and the revised method[10]. Their major finding was that rates measured using the dissolution method were on average 1.7 times greater than those derived from the bubble method. Accordingly, the authors used this number in a first-pass calculation to adjust estimates of global nitrogen-fixation rates upward.

A secondary finding of this work[2] was that there was a "poor overall correlation between the N_2-fixation rates calculated using the two methods". The magnitude of the difference between the methods varies, and in a few instances the methods actually generated similar values. Großkopf *et al.* discuss mechanisms that may explain this variation. Their explanations include the use of different time intervals between the injection of a $^{15}N_2$ bubble into a sample and the onset of the daily period of biological N_2 fixation, as well as the variable buoyancy of diazotrophs and hence their proximity to the $^{15}N_2$ bubble in the experiments. Intriguingly, both of these mechanisms seem to reflect the structure of diazotroph communities, to the extent that the identity of the dominant diazotroph along any cruise transect will affect the N_2-fixation rates measured using the bubble-addition technique (Fig. 1).

Großkopf and colleagues' findings will surely stimulate a spirited and necessary debate on how, and if, we can reinterpret the past decade or so of N_2-fixation measurements. In the past, the bubble-addition assay has been implemented inconsistently — different researchers used a range of incubation durations; experiments were initiated at different times of day; varying volumes of $^{15}N_2$ were injected; and experiments unavoidably targeted a spectrum of nitrogen-fixing communities. Given Großkopf and colleagues' results and this legacy of methodological inconsistency, some past measurements may need no revision, whereas others probably need large adjustments. Unfortunately, reliably identifying which measurements require revision may not be possible. The way forward is therefore clear: to understand the current state of the oceanic nitrogen cycle, researchers need to forge a path to a standardized methodology and to assess whether the database of historical N_2-fixation rates[3] can be adjusted, or whether field efforts will need to begin anew. ■

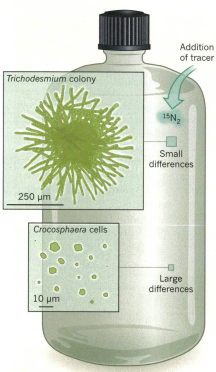


Figure 1 | Marine nitrogen fixation. A commonly used method for measuring the rate of nitrogen (N_2) fixation by diazotrophic microorganisms in the ocean is to add isotopically labelled nitrogen gas, $^{15}N_2$, to a sample of sea water, and then to measure the uptake of this nitrogen tracer by diazotrophs in the sample. The $^{15}N_2$ has classically been added as a bubble of gas[4], but Großkopf *et al.*[2] have compared the results of this traditional approach with those of a modified, more reliable version of the assay[10] — in which the tracer is added as a solution in filtered sea water — for measurements of the N_2-fixation rates of diazotrophs in the Atlantic Ocean. They found that the rates obtained using the modified method were, on average, greater than those obtained using the traditional approach. The differences between rates may in part track the composition of the diazotroph communities sampled — the discrepancies were larger when unicellular cyanobacteria, such as *Crocosphaera*, were in the majority, and smaller when the colony-forming genus *Trichodesmium* predominated.

Angelique E. White *is at the College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA.*
e-mail: awhite@coas.oregonstate.edu

1. Codispoti, L. A. *Biogeosci. Discuss.* **3,** 1203–1246 (2006).
2. Großkopf, T. *et al. Nature* **488,** 361–364 (2012).
3. Luo, Y.-W. *et al. Earth Syst. Sci. Data Discuss.* **5,** 47–106 (2012).
4. Montoya, J. P., Voss, M., Kahler, P. & Capone, D. G. *Appl. Environ. Microbiol.* **62,** 986–993 (1996).
5. Mahaffey, C., Michaels, A. F. & Capone, D. G. *Am. J. Sci.* **305,** 546–595 (2005).
6. Zehr, J. P., Mellon, M. T. & Zani, S. *Appl. Environ. Microbiol.* **64,** 3444–3450 (1998).
7. Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N. & Dunne, J. P. *Nature* **445,** 163–167 (2007).
8. Zehr, J. P. *et al. Science* **322,** 1110–1112 (2008).
9. Foster, R. A. & O'Mullen, G. D. in *Nitrogen in the Marine Environment* 2nd edn (eds Capone, D. G. *et al.*) 1197–1218 (Elsevier, 2008).
10. Mohr, W., Großkopf, T., Wallace, D. W. R. & LaRoche, J. *PLoS ONE* **5,** e12583 (2010).

nature

16 August 2012 / Vol 488 / Issue No. 7411

Publication policies, author guidelines and information

This is a summary of *Nature*'s guide to authors, publication policies and other author information, available in full on the web via our main author information page at www.nature.com/nature/authors.

Getting published in *Nature*

Please read www.nature.com/nature/authors/get_published before finalizing any original scientific research manuscripts you intend to submit. This section explains *Nature*'s editorial criteria, and how manuscripts are handled by the editors after submission and before acceptance for publication.

Formatting guide

Please read www.nature.com/nature/authors/gta before finalizing and submitting your contribution. This section provides instructions for preparing, formatting and writing all types of contribution published by *Nature*.

Publication policies

This section, at www.nature.com/authors, describes *Nature*'s conditions of publication, including policies on authorship; competing financial interests; duplicate publication; prepublicity and embargoes; materials and data availability; research on animal (including human) subjects; referees; corrections, retractions, errata and addenda to published papers; the complaints procedure and refutations; and copyright/licence to publish. All contributions, for whatever section of the journal, are considered for publication and published on the condition that authors agree to these policies.

Submissions

These sections, at www.nature.com/nature/authors/submissions, contain our online service for presubmission enquiries and submission and tracking of Articles, Letters and Review-type manuscripts. Authors should read the formatting guide before using this service. Authors who cannot submit online can submit by post, but *Nature* strongly recommends the online service. **Authors must not submit by both methods.** Authors experiencing operational difficulty in using the online submission service can send an enquiry to mtshelp@nature.com, with a brief explanation of the problem.

Editorial contacts and author services

This section, at www.nature.com/nature/about/contact, provides the addresses of *Nature*'s editorial offices; names of editorial staff; and contact details for advice, reprints, permissions, press office (press@nature.com), and other author-related contact information, including postal addresses for mailed submissions. Short biographies of editors are also available at www.nature.com/nature/about/editors.

Final submissions

This section, at www.nature.com/nature/authors/submissions/final, applies only if an Article or Letter has been accepted in principle for publication and the author has been asked by the editor for a final version. Authors will need to upload electronic files of the text, production-quality figures and supplementary information (if any) to our database. Authors must include a completed manuscript checklist and all relevant signed forms and declarations as requested by the editor.

Short version of guide to authors

See www.nature.com/nature/authors for full version. Contributors not used to *Nature*'s format are advised to read some papers in published issues before submission, to familiarize themselves with *Nature*'s style. *Nature*'s main formats for original research are Articles and Letters.

ARTICLES are original reports whose conclusions represent a substantial advance in understanding of an important problem and have immediate, far-reaching implications. They do not normally exceed **5 pages** of *Nature* and have **no more than 50** references. (One page of undiluted text is about 1,300 words.)

Articles have a summary for readers outside the discipline, separate from the main text and containing no references, of up to 150 words. The summary does not contain numbers, abbreviations, acronyms or measurements unless essential. It contains a brief account of the background and rationale of the work, followed by a statement of the main conclusions, introduced by the phrase 'Here we show' or its equivalent.

The Article itself typically has about 3,000 words of text, beginning with up to 500 words of referenced text expanding on the background to the work, followed by a concise, focused account of the findings, and ending with one or two short paragraphs of discussion. The text can contain a few short subheadings consisting of about 20 but no more than 40 characters (including spaces) each.

LETTERS are short reports of original research focused on an outstanding finding whose importance means that it will be of interest to scientists in other fields. They do not normally exceed **4 pages** of *Nature*, and have **no more than 30** references. They begin with a **fully referenced paragraph** of not more than 300 words, aimed at readers in other disciplines. This paragraph contains a summary of the background and rationale for the work, followed by a one-sentence statement of the main conclusions introduced by the phrase 'Here we show' or its equivalent, and ending with two to three sentences describing the implications in context.

The main text which follows is typically up to 1,500 words long. This text can, if the author requires, start with a further brief paragraph of introductory material. A description of the research then follows. Any discussion at the end of the text should be as succinct as possible.

FIGURES in both Articles and Letters are as small as possible: Articles typically have 5 or 6 figures, Letters typically have 3 or 4 figures. Legends and/or methods are additional to the main text. Methods Summaries (which are published in print only) are no longer than 300 words; Methods sections (which are published online) can be up to 1,000 words; legends should not exceed 100 words each and start with a one-sentence title for the figure, followed by a brief explanation of the parts and symbols.

See www.nature.com/nature/authors/gta/others.html for details of other types of contribution to *Nature*, including Brief Communications Arising; Reviews and Perspectives (presubmission enquiries to reviewsandinsights@nature.com); and Correspondence (submit to corres@nature.com).

See Authors and Referees @ NPG (www.nature.com/authors) for details of Nature journals' publication policies, author and referee services, author benefits and for links to Nautilus and Peer to Peer blogs.

nature INSIGHT

CHEMISTRY AND ENERGY

Taking our solar energy commitment to the next level too

Active in solar energy for more than 25 years, Total has taken its technological and manufacturing capabilities to a new level with the acquisition of an interest in SunPower*. Our goal is to develop a reliable, efficient and competitive solar energy business and become a global leader in this segment.

www.total.com

* SunPower Corp. designs, manufactures and supplies most efficient solar modules and panels on the market.

nature

nature

A key thread in your scientific career

Whether in print, online or mobile, *Nature* is always at the forefront of science and is your forum to read, watch, listen and engage with key research, news and opinion.

How will you access the next landmark discovery?

NATURE.COM/NATURE

nature publishing group npg

nature INSIGHT

CHEMISTRY AND ENERGY

16 August 2012 / Vol 488 / Issue No 7411

Cover illustration:
S. S. Moran/
iStockphoto

Editor, *Nature*
Philip Campbell

Publishing
Nick Campbell
Claudia Deasy

Insights Editor
Karl Ziemelis

Production Editor
Jenny Rooke

Art Editor
Nik Spencer

Sponsorship
Reya Silao
Yvette Smith
Gerard Preston

Production
Emilia Orviss

Marketing
Elena Woodstock
Hannah Phipps

Editorial Assistant
Rina Nozawa

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7833 4000
e: nature@nature.com

nature publishing group

Life, as many people know it, would not be possible without access to cheap and abundant energy. More than 85% of the world's energy comes from the combustion of oil, coal and natural gas. Scientists generally agree that the combustion of fossil fuels is causing warming of Earth's climate and having deleterious effects on the environment. Many fear that climate change will lead to coastal flooding, an increase in the prevalence of vector- and water-borne infectious diseases, and conflict over fossil fuels, water and food. Clearly, action is needed.

Reducing our consumption of energy and adopting a more frugal lifestyle could be the first steps, but they cannot be the only strategies. The fossil fuels that remain must be used more effectively and other energy sources need to be identified, preferably ones that are climate neutral. The Reviews in this Insight cover a small fraction of the various scientific and engineering endeavours that may eventually deliver the technological innovations to harvest energy from alternative sources.

The Reviews clearly show that there are many energy solutions to be explored. These possible strategies range from the well-established photovoltaics to nascent methods, such as the use of algae to produce potential fuels. However, we still have a long way to go before approaches such as those covered in this Insight can compete in earnest with conventional energy sources. We hope that the articles that appear in the following pages will inspire young scientists and engineers to join the quest to secure a sustainable-energy future.

We are pleased to acknowledge the support of TOTAL in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer review.

Rosamund Daw, Joshua Finkelstein & Magdalena Helmer
Senior Editors

CONTENTS

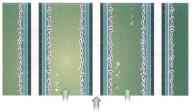
PERSPECTIVE

294 **Opportunities and challenges for a sustainable energy future**
Steven Chu & Arun Majumdar

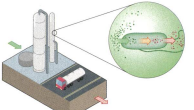
REVIEWS

304 **Materials interface engineering for solution-processed photovoltaics**
Michael Graetzel, René A. J. Janssen, David B. Mitzi & Edward H. Sargent

313 **Membrane-based processes for sustainable power generation using water**
Bruce E. Logan & Menachem Elimelech



320 **Microbial engineering for the production of advanced biofuels**
Pamela P. Peralta-Yahya, Fuzhong Zhang, Stephen B. del Cardayre & Jay D. Keasling



329 **Exploiting diversity and synthetic biology for the production of algal biofuels**
D. Ryan Georgianna & Stephen P. Mayfield

PERSPECTIVE

doi:10.1038/nature11475

Opportunities and challenges for a sustainable energy future

Steven Chu[1] & Arun Majumdar[1]

Access to clean, affordable and reliable energy has been a cornerstone of the world's increasing prosperity and economic growth since the beginning of the industrial revolution. Our use of energy in the twenty–first century must also be sustainable. Solar and water–based energy generation, and engineering of microbes to produce biofuels are a few examples of the alternatives. This Perspective puts these opportunities into a larger context by relating them to a number of aspects in the transportation and electricity generation sectors. It also provides a snapshot of the current energy landscape and discusses several research and development opportunities and pathways that could lead to a prosperous, sustainable and secure energy future for the world.

The industrial revolution began in the mid-eighteenth century, and provided humans with capabilities well beyond animal and human power. Steam-powered trains and ships, and then internal combustion engines transformed how people moved and produced goods around the world. Electrification and related technologies continued the revolution in the nineteenth and twentieth centuries. Today, a growing number of people keep their homes warm in the winter, cool in the summer and lit at night. They go to the local market in cars with the power of over a hundred horses and fly across continents in wide-body aeroplanes with the power of a hundred thousand horses. This power is derived largely from our ability to exploit fossil sources of energy. However, in the transition from human and horse power to horsepower, the carbon emissions that result from the equivalent of over a billion horses working continuously have created significant climate-change risks.

At the beginning of the industrial revolution, the population of the world was 700 million. Today, the population is 7 billion and is estimated to grow to 9 billion by 2050, and about 10 billion by 2100 (ref. 1). Most of this population growth will be in Asia and Africa, where rapidly rising economic growth will place additional demands on energy supply. The International Energy Agency (IEA) based in Paris has projected that the world's energy demand will increase from about 12 billion tonne oil equivalents (t.o.e.) in 2009 to either 18 billion t.o.e. or 17 billion t.o.e. by 2035 under their 'current policies' or 'new policies' scenarios, respectively[2]. Carbon-dioxide emissions are expected to increase from 29 gigatonnes per year to 43 Gt yr^{-1} or 36 Gt yr^{-1} under the current and new policies, respectively. The actual path we follow will depend on how efficiently and effectively we use existing and new sources of energy.

The world needs another industrial revolution in which our sources of energy are affordable, accessible and sustainable. Energy efficiency and conservation, as well as decarbonizing our energy sources, are essential to this revolution. Reducing carbon emissions on the timescale needed to mitigate the worst risks of climate change will not be driven by our inability to find cost-effective sources of fossil fuels. Despite the significant growth in the use of renewable energy, the fractional sum of non-carbon- emitting sources of energy that remained constant during the past two decades is sobering (Fig. 1).

Energy systems can be divided into transportation and stationary. The supply, demand and distribution infrastructures within each system are highly coupled, but are currently largely independent of each other. In this Insight, the Reviews focus on specific subareas of biofuels, solar electricity generation, and electricity generation from salinity gradients and waste water. The systems discussed are largely at the research and development stage, and are not yet ready to displace more mature sources of energy. Nevertheless, we should be constantly looking for innovations that make a marked improvement on today's approaches or provide an entirely new approach. To place the topics discussed in the Reviews in context, this Perspective discusses transportation and stationary power systems, and indicates how the topics discussed in the Reviews fit into the energy landscape.

Many important and challenging research areas have the potential to significantly affect our future energy needs. For example, energy efficiency and the integration of energy sources with electricity transmission, distribution and storage are vitally important, but these are only briefly mentioned in this Perspective. Excellent reviews of these topics include the National Academies' report *America's Energy Future*[3] and the US Department of Energy's (DOE's) *Report on the First Quadrennial Technology Review*[4].

Transportation

Petroleum-derived liquid fuels are the overwhelming source of energy in the current transportation infrastructure. The geographical distribution of petroleum resources is changing as reserves are found and accessed with improved technologies for discovery and production. However, this distribution of oil supply generally does not coincide with where the demand is located. For example, many countries import oil at an unprecedented scale, which can lead to significant balance-of-trade and national-security challenges. In 2011, about 2.690 billion tonnes of oil were consumed; of this, 1.895 billion tonnes of crude oil and 0.791 billion tonnes of refined products crossed national borders[5] and significant discoveries of oil, natural-gas liquids and natural gas could potentially alter the global-energy landscape[6].

We will review some of the opportunities and challenges related to transportation technologies. For example, most of the future infrastructure of the world will be built in locations where we have the greatest opportunity to transition to sustainable mobility. Desirable and affordable public transportation that is fully integrated into urban planning[7], and the use of information technologies to assist and displace transportation can significantly reduce fuel consumption, but are not discussed in this Perspective.

The Quadrennial Technology Review (QTR) of the US DOE provides

[1]US Department of Energy, 1000 Independence Avenue SW, Washington DC 20585, USA.

a broad overview of state-of-the-art technologies and opportunities for future research[8]. Improvements in energy efficiency of vehicles can greatly reduce oil dependency. These improvements include increased use of light-weight materials, such as advanced ultra-high tensile strength steels, aluminium and magnesium alloys, polymers, and carbon-fibre reinforced composite materials[9]. The integration of lighter weight materials is especially important if more complex parts can be manufactured as a single unit. The potential for reducing the weight of vehicles has already been shown, and in the next 10–20 years, an additional 20–40% reduction in overall weight, without sacrificing safety, seems to be possible[10]. For every 10% weight reduction of the vehicle, an improvement in fuel consumption of 6–8% is expected[11].

Reducing energy losses as a result of friction is also possible[12] (Fig. 2). Advances in cost-effective technologies such as tribology, tyres, braking and waste-heat energy recovery, and aerodynamics could potentially lead to efficiency improvements of 20% in the short term and more than 60% over a longer term (15–25 years).

Internal combustion engines

The internal combustion engine using liquid-transportation fuel (or liquid ICE) will probably continue to have a major role over the next few decades. However, improvements to the efficiency of the liquid ICE are possible because the efficiency of most spark-ignition engines is typically 25–35%, whereas that for compression-ignition diesel engines is about 40–50%. With direct injection, lean burn and turbocharger technologies, the spark-ignition ICEs operating on premium octane ratings can approach diesel efficiencies. The combination of *in situ* measurements of prototype ICE designs with detailed simulations, which are made possible with high-performance computers, are increasing the energy efficiency and have lowered emissions[13,14]. Finally, low-cost waste-heat recovery can increase efficiency, especially in heavy-duty vehicles. Approaches include use of the Rankine cycle to convert waste heat to work, and the development of low-cost and high-efficiency solid-state thermoelectric systems[15].

Battery-based electrification

Plug-in hybrid and all-electric light- and medium-duty vehicles have the opportunity to displace a significant amount of liquid fuel use in transportation. The main challenges are performance and cost of the battery systems. The performance of battery systems is quantified by usable energy density, power density (including fast charging), cycle lifetime and robustness. In the past 5–6 years, a remarkable amount of progress in research has been made in battery cathodes, anodes and electrolytes. Large volumetric changes in the electrodes have led to designs in which micro- or nanostructures are embedded in a conducting and flexible matrix to allow for the relief of mechanical stresses. State-of-the-art batteries based on graphite anodes and lithium-manganese-oxide composite cathodes for lithium batteries are being commercialized[16]. Within the next few years, battery system packs of 200 Watt-hour kg^{-1} at a charging rate of full charge in 3 hours, which is double the current cell-energy density, are expected to become available. The current production cost of a vehicle battery is estimated to be US$650 kWh^{-1} of usable energy storage, but this is expected to drop to below $150 kWh^{-1} by 2030 (ref. 17).

The US DOE is supporting a research and development effort that would allow a four or five passenger plug-in hybrid electric vehicle (PHEV) or an electric vehicle with about a 160-km range to be competitive as a mass-market ICE car within a decade. The US DOE EV-Everywhere challenge for PHEV, will require the cost of the battery system to be reduced by an estimated $190–300 kWh^{-1}, depending on whether the car is a PHEV or an electric vehicle. The development of anode-protecting materials and non-flammable electrolytes that are stable at high voltage and tolerate 55 °C are desired characteristics of third-generation lithium-ion batteries. Developing batteries beyond lithium-ion batteries, such as lithium-sulphur and metal-air batteries, could achieve up to ten times the energy density of the current lithium-ion batteries, but materials research is needed to develop anode and cathode protection, as well as non-flammable electrolytes with electrochemical stability over a large potential range.

Battery packs typically use about 50% of the total battery capacity, and the charging rates are limited to increase the lifetime[18]. If sensor technologies are developed that can continuously monitor the properties of individual cells, such as internal impedance, temperature and

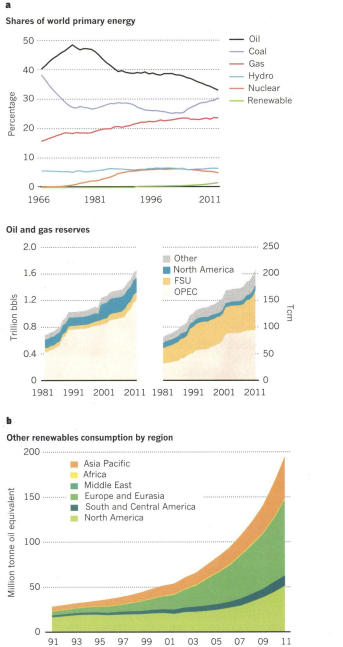


Figure 1 | Statistical review of world energy. **a**, Fossil energy comprises roughly 86% of the world's main energy consumption. Although the consumption of oil has increased by 31% between 1980 and 2008, the known reserves have increased comparably owing to improvements in exploration and extraction technologies. Much of the world's shale-gas reserves are not 'proven' and are not included. The fractional sum of non-carbon emitting sources of energy remained constant during the same time period. **b**, Growth of renewable energy was offset by the decline in nuclear power generation. Renewable energy sources in power generation grew by 17.7%. Wind generation (25.8%) accounted for more than half of renewable power generation for the first time. Renewables accounted for 3.8% of global power generation, with the highest share in Europe and Eurasia (7.1%). Adapted with permission from ref. 72.

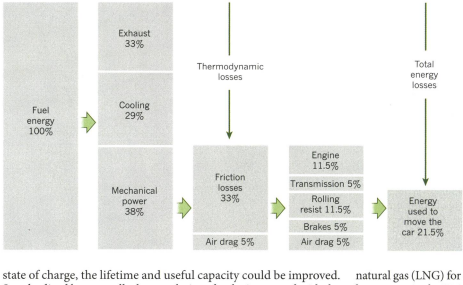


Figure 2 | Vehicle energy losses. Of the energy that fuel provides to vehicles a substantial proportion is lost. A breakdown of the average losses of internal-combustion-engine cars (fleet make up 70% petrol and 30% diesel) is shown. Heat lost constitutes 30–37% of the energy as a result of exhaust gases with lower energy content and convection. The other losses come from heat dissipation (25–33%), mechanical losses (33–40%), air drag (3–12%), rolling friction (12–45%) and brake losses (about 5%). These losses mean only about 21.5% of the energy is used to move the car. Adapted with permission from ref. 12.

state of charge, the lifetime and useful capacity could be improved. Standardized battery cells that are designed to be integrated with the original equipment manufacturer's thermal management systems could also reduce cost[19].

Fuel-cell-based electrification

The high efficiency of fuel-cell-powered electric vehicles makes this form of electrification a potentially viable option for the future. Investment in this technology is driven by the potential of extended range and faster fuelling times of moderately low-priced cars. Fuel-cell cost has been lowered and their lifetime increased[20], but further gains are needed. Platinum-group catalyst loading has been reduced fivefold since 2005; however, further reductions are needed[4,21], or these catalysts need to be replaced with less costly alternatives. Costs can also be lowered and performance improved through robust higher-conductivity and higher-temperature membranes, improvements in balance-of-plant components, such as humidifiers and compressors, as well as thermo-fluid design and control.

There are inherent volumetric energy density issues for hydrogen-gas storage. To achieve a range of 480 km, fuel-cell electric vehicles need to store about 4–7 kg of hydrogen[4,22]. A carbon-fibre-composite tank pressurized to 700 bar is the best current option for personal vehicles, but this costs[23] about $3,000. Research is under way to develop materials and manufacturing processes to reduce the cost of composite tanks. In parallel, researchers are searching for lower-pressure storage assisted by high-surface area materials that could physisorb or weakly chemisorb hydrogen and still maintain fast-fuelling times[24,25].

The supply infrastructure and a low-carbon source of hydrogen are also a challenge. The technology advances in shale-gas production, and the possibility of large reserves in Europe and Asia, in addition to the considerable reserves in North America, could have a significant affect on the transportation sector. In addition to the direct use of natural gas as a fuel (see later), low-cost natural gas could spur the deployment of local reforming or hydrogen filling stations for near-term hydrogen production. Alternatively, commercial reforming plants, such as hybrid power plants that produce hydrogen as well as CO_2, for enhanced oil recovery located near oil-field and refinery sites can serve as an economical source of hydrogen[4,26]. In regions that are close to a large commercial production plant, delivery to local filling stations can be made through high-pressure gaseous tube trailers, but in the long run, a cost-competitive method to produce hydrogen with considerably lower net carbon emissions is needed.

Natural gas for transportation

The projected low cost of natural gas in the United States in the next few decades compared with that of petrol is expected to lead to wider adoption of natural-gas vehicles. Displacing diesel fuel with liquefied natural gas (LNG) for class 8 tractor-trailer trucks commonly used on long routes in the United States is already economically viable because a typical long-haul truck uses about 90,000 litres of fuel per year (about $80,000 per year in fuel costs, today). The incremental purchase price of LNG trucks can be up to $100,000 per truck for cryogenic tanks and related upgrades, as a result of low-volume market conditions. Even so, the payback period is currently 3–4 years on a net-present-value basis using a 7% discount rate, and would drop considerably with even modest increases in production volumes[27]. Developing a fuelling infrastructure for LNG long-haul trucks would require fuelling stations about every 240–320 km for a truck range of 800–960 km. Commercial viability is reinforced by plans in the private sector to make the required investments in infrastructure. LNG-powered freight trains are also being considered.

Compressed natural gas (CNG) has been used for buses, delivery trucks and light-duty vehicles. To make vehicles using CNG economically viable without subsidies, low-cost CNG storage technologies are needed. In the United States, light-duty vehicles account for 75% of on-road fuel consumption. There are roughly 160,000 gasoline (petrol) service stations in the United States[28]; creating a similar nationwide infrastructure for CNG vehicles would be prohibitively expensive (more than $100 billion). However, about 60 million homes in the United States have natural-gas delivery. Economic viability for CNG cars and refuelling systems can be achieved if the payback period for the additional system-level cost is typically 5 years or less; at present it is about 10–15 years for a vehicle with average mileage. For vehicles with high annual mileage and for vehicles with low kilometres per litre, natural-gas vehicles can have less than a 5-year payback, even today. Research into fibre-matrix composites for high-pressure light-weight tank materials is needed, as well as into natural-gas sorbents for low-pressure storage. Although seldom discussed, multifuel ICEs can be designed to operate on CNG for 30–60 km (the CNG–ICE hybrid equivalent of a PHEV) and then switch to petrol. Similar to a PHEV, a CNG–petrol–ICE vehicle could compensate for the partial coverage of CNG fuelling stations.

Natural gas can also be converted into liquid fuels using either the Fischer–Tropsch or the methanol process. The capital cost per barrel of liquid fuel reduces with increasing capacity of a Fischer–Tropsch plant according to scaling laws. However, with increasing capacity, capital costs are suggested to deviate and are higher than scaling law predictions, which increases the financial risk for gas-to-liquid plants. Research is needed to find alternative approaches for exciting the carbon–hydrogen bond and synthesizing carbon–carbon bonds. Biological approaches that use organisms, such as methanotrophs, that can metabolize natural gas and produce long-chain hydrocarbons seem worth exploring. Even if this approach is successful at laboratory scales, it will need to be scalable to large-volume production. Large quantities of methanol are already produced from natural gas for industrial

purposes at costs that are roughly equivalent to petrol. Methanol could be used in a petrol–alcohol blend, much like ethanol in the United States. However, pure or high-percentage methanol-based transportation could face distribution-infrastructure challenges.

Alternative liquid-transportation fuels

Liquid fuels derived from oil became the main form of energy for transportation largely because of their high energy densities. Associated with their high energy content, the energy transfer rate during vehicle refuelling is about 6 MW; in contrast, electrical charging will be tens of kilowatts. Apart from inherent limitations in battery chemistries, there are practical limits to the size of the electrical connector that could accommodate megawatt-scale power transfer. However, a search for alternatives to oil for transportation energy is required to deal with the growing concerns over the rising and volatile price of oil, the vulnerability to supply disruptions, and balance-of-trade issues. Biofuels, particularly those produced domestically at competitive prices, would strengthen a nation's energy and economic security.

There are a number of approaches to alternative transportation fuels being actively explored (Fig. 3). The estimated future (10–15 year) costs for classes of alternative fuels can vary[29] (Fig. 4). The US Environmental Protection Agency (EPA) estimates that Brazilian sugarcane ethanol — already price competitive with oil-based fuels in Brazil — reduces total life-cycle greenhouse gas emissions, including direct and indirect land-use change emissions, by 61% (ref. 30). There are numerous, and sometimes contentious, studies of carbon life-cycle emissions[30–32]. Minimization of indirect land-use concerns and the sequestration of process CO_2 could result in net-negative carbon emissions, so that fuel production and use becomes a net carbon sink. On the other hand, if biofuels based on energy-intensive crops were coupled with poor-land management this could result in environmental costs that are far higher than those associated with oil-based fuels[29].

The production of ethanol by domesticated yeast for fermentation is perhaps 4,000 years old. In this Insight, Paralta-Yahya *et al.*[33] review the application of metabolic engineering and synthetic biology to alter microbes for the production of advanced biofuels, and of precursors to drop-in substitutes for petrol, diesel and jet fuels. Static adjustment of transcription, translation and post-translational modifications does not allow engineered organisms to respond to changing bioreactor conditions and cellular changes. Remarkably, dynamic sensing and regulation of fatty-acid ethyl ester (FAEE) intermediates in *Escherichia coli* has been shown to increase FAEE yield by threefold, to reduce the cellular concentration of toxic intermediates and to significantly improve the genetic stability of producing strains.

Successful commercial-scale deployment of this class of technologies will depend on microbial productivity and robustness. The cost of microbial feedstock is also a major factor: production of advanced fuels from simple sugars and starches is further along in the development of these fuels, but it is held back by high feedstock prices. Lignocelluloses, including agricultural and wood-waste streams, are roughly an order of magnitude less costly, and are generally viewed as the end-goal feedstock. Much attention is now focussed on reducing the cost of converting lignocelluloses into forms that are more readily used by microbial organisms.

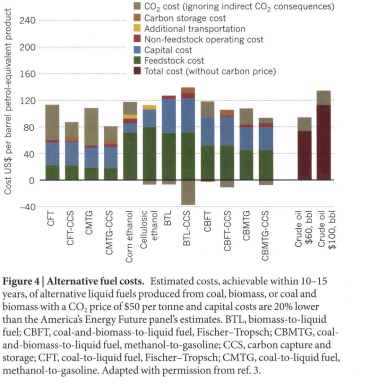


Figure 4 | Alternative fuel costs. Estimated costs, achievable within 10–15 years, of alternative liquid fuels produced from coal, biomass, or coal and biomass with a CO_2 price of $50 per tonne and capital costs are 20% lower than the America's Energy Future panel's estimates. BTL, biomass-to-liquid fuel; CBFT, coal-and-biomass-to-liquid fuel, Fischer–Tropsch; CBMTG, coal-and-biomass-to-liquid fuel, methanol-to-gasoline; CCS, carbon capture and storage; CFT, coal-to-liquid fuel, Fischer–Tropsch; CMTG, coal-to-liquid fuel, methanol-to-gasoline. Adapted with permission from ref. 3.

The net energy yield per hectare per year is a major factor in determining feedstock cost. The overall yield for biofuels varies widely: Brazilian sugarcane can produce as much as 800 gallons per acre-year, which is about twice as high as productivities based on North American maize. High-yield grasses or fast-growing trees that could be feedstocks for advanced biofuels have the potential to surpass sugarcane productivities by 1.3 times or more[34]. Most lignocellulosic feedstocks have low-volume energy density: relative to fossil fuels they are 'light and fluffy'. To improve the economies of scale of biorefineries, especially thermochemical refineries, increasing the feedstock collection radius is desirable. Biofuels could become more competitive if a means of efficiently concentrating the biomass during the harvesting process (such as combining the reaping and pelleting the woody materials) could be developed. The same densification of the feedstock could also lead to the biofuel equivalent of grain elevators. Reducing 'first touch' costs are an integral part of efficient agricultural practice, and progress in this area would significantly lower the overall feedstock cost delivered to biorefineries.

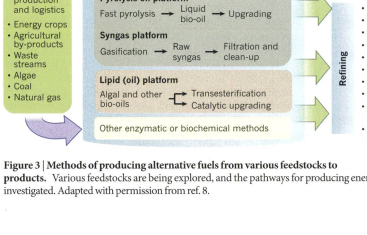


Figure 3 | Methods of producing alternative fuels from various feedstocks to products. Various feedstocks are being explored, and the pathways for producing energy or fuel investigated. Adapted with permission from ref. 8.

In this Insight, Georgianna and Mayfield[35] discuss the use of algae to produce next generation biofuels. Algae can provide high fuel yields, especially if areal sunlight collection can be used to drive volumetric growth of algae. However, algal growth in closed systems requires very high capital investments relative to energy crops such as grasses or trees. In open-pond systems, water use is a major issue but may be partially ameliorated through the use of species that can grow in brackish or saltwater.

Other approaches to bio-based fuel production include

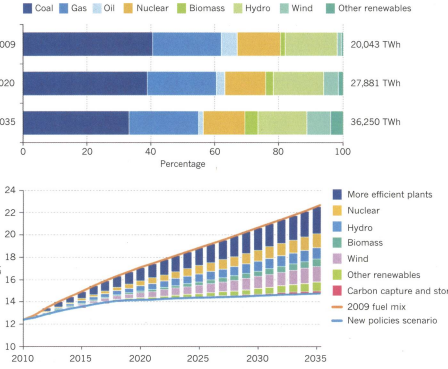


Figure 5 | Projections from the International Energy Agency for the generation of electricity until 2035. In the top panel, the relative energy mix in 2009 and that projected in the International Energy Agency's 'new policies' scenario. In the bottom panel, in the new policies scenario, carbon emissions continue to grow, but slower than the 2009 fuel-mix curve. The 2009 fuel-mix curve is based on emissions that would have been generated for the projected level of electricity generation were there no change in the mix of fuels and technologies, and no change in the efficiency of thermal generation plants after 2009. The cost of renewable energy is declining at a faster rate than assumed in the International Energy Agency's report. In addition, the relative mix of natural gas to coal may be larger than that projected. Adapted with permission from ref. 38.

the manipulation of photosynthetic bacteria to produce biofuels, diverting high-energy Calvin-cycle intermediates upstream of glucose — for example energy-dense terpene (a biofuel precursor) production in trees or microbes — or developing alternatives to C_1 and C_4 carbon fixation.

Biological enzymes can synthesize carbon–carbon bonds with an unparalleled high specificity, but photosynthesis may not be the only approach to converting sunlight into hydrocarbon fuels[36]. Several non-photosynthetic alternatives are under investigation that could potentially overcome many of the limitations of photosynthesis. To appreciate the potential of such approaches, the energy conversion process can be divided into three steps. First is to identify what reducing equivalents, other than solar photons captured through photosystems I and II, can be accepted by various microorganisms. Several organisms are known to be capable of growth on hydrogen sulphide, hydrogen, electrons, ammonia and reduced ions such as iron (II). Second is to investigate opportunities to fix CO_2 using pathways other than those used in C_3 or C_4 plants. Potential systems may include the reverse tricarboxylic-acid cycle (often called the reverse Krebs cycle), the Woods–Ljungdahl cycle used by acetogens, the hydroxypropionate–hydroxybutyrate cycle or newly designed biochemical pathways. The final step is to determine whether we can metabolically engineer direct carbon products into a molecule such as acetyl-CoA, which is a precursor for many energy-dense fuels. These three steps can be engineered into autotrophic organisms, an approach that is now being supported by a US DOE Advanced Research Projects Agency-Energy programme called Electrofuels[37].

Finally, researchers are investigating highly efficient non-biological energy-conversion approaches that generate fuel from sunlight by the oxidation of water into hydrogen and oxygen and reduction of CO_2 to fuel. The Joint Center for Artificial Photosynthesis, a US DOE funded Energy Innovation Research Center, was established to identify Earth-abundant, robust light absorbers with optimal bandgaps to harvest sunlight most effectively and efficiently, to accelerate the rate of catalyst discovery for solar energy-to-fuel conversion reactions and to provide system integration and scale-up so that laboratory experiments can quickly transition into prototypes for commercial development.

Clean and affordable electricity generation

The IEA has projected, to 2035, how electricity is expected to be generated[38] (Fig. 5). In the agency's new policies scenario, the growth of carbon emissions is curtailed, but does not decline. The scenario assumes that government policies such as carbon pricing are adopted in several Organisation for Economic Co-operation and Development countries and in China, but the projections of cost reduction in renewable energy are overly conservative. As an example, the agency estimates the full levellized cost of electricity (LCOE) of onshore wind to be about \$90 MWh^{-1} in 2010 real dollars by 2020 (ref. 39). Similarly, the US Energy Information Administration estimates[40] the LCOE of wind for 2016 will be about \$80–120 MWh^{-1} (Fig. 6).

In contrast to these estimates, an analysis of the LCOE of various forms of renewable energy, based on the available data from projects in which investment or purchase contracts have been completed[41], shows that wind at level (class) 4 sites in the United States is currently about \$73 MWh^{-1} (Fig. 7). These costs do not include transmission line costs, which average[42] at about \$300 kW^{-1}. Some experts estimate that by 2020, the LCOE of wind energy at level 4 sites will fall below \$60 MWh^{-1}. In the United States, utility-scale solar photovoltaic projects are being installed in 2011–2012 at an unsubsidized cost of about \$150 MWh^{-1} (the US investment tax credit has been taken out[43,44]), which is consistent with the Bloomberg New Energy Finance estimates[41]. Utility-scale solar photovoltaics are projected to be between \$60–120 MWh^{-1} in good insolation areas. The lowest LCOE for combined-cycle natural gas is about \$50–60 MWh^{-1} in the United States, assuming gas costs of \$3–4 per million British thermal units. Although LCOE is an important factor, time of delivery, LCOE at various size scales and the potential for energy storage are also important.

Affordable and environmentally sustainable electricity generation is already at cost parity in areas in which energy costs exceed \$200 MWh^{-1}. On the basis of the current and projected learning curves for wind and solar energy, renewable energy will become increasingly affordable on a global scale. In addition, energy storage with durable and inexpensive batteries will allow electricity accessibility in micro- or meso-scale grids to leapfrog the need to bring centrally generated electricity to distant rural areas.

Even though the costs of renewable energies such as wind and solar continue to decline, the full economic benefit of these variable sources of energy will not be realized until the more established forms of energy generation are integrated with renewable sources into transmission and distribution, load response and storage of electricity. Eighty per cent

penetration of renewable energy in the United States is possible by 2050, but to achieve even half this goal, a transformation of the electricity system on the basis of technology advances and new operating procedures, business models and regulatory approaches is required[45].

Solar

In 2010, the US DOE created the SunShot initiative to reduce the LCOE from solar sources (photovoltaic and thermal) to $50–60 MWh^{-1}, without subsidies, by 2020. For solar photovoltaics at utility scale, the fully installed cost target must be about $1 per Watt of electricity generated. The costs are broken down into $0.50 W^{-1} for the solar module, $0.10 W^{-1} for the power electronics for grid integration and $0.40 W^{-1} for the balance of the system (including installation, permits and all other costs)[46].

In the four years between the second quarter of 2008 and the second quarter of 2012, the actual spot market price of silicon solar modules for immediate delivery declined from $4 W^{-1} to about $1 W^{-1} (ref. 47). Although part of this drop was due to a temporary oversupply coupled to the worldwide recession, continual engineering and manufacturing innovations contributed to this decline and are likely to continue doing so in the coming years: future delivery contracts are about $0.80 W^{-1}, and there is a growing confidence that module costs will achieve the goal of $0.50 W^{-1} by 2020 (ref. 48). In contrast, the balance-of-system costs, specifically the soft costs such as permits, interconnection and inspection, financing and customer acquisition have declined but at a significantly slower rate. Because of this trend, a focus of the US DOE SunShot programme is lowering all balance-of-system costs commensurate with module prices[49].

All costs are affected by increasing the efficiency of the solar cells while reducing their production cost. The Shockley–Queisser limit is a well-known efficiency limit of single-junction solar cells, which suffer from spectrum losses, recombination and black-body radiation. The Shockley–Queisser limit of silicon is 32.7%, but a practical limit is considered to be 29% (ref. 50). The ultimate efficiency limit in the conversion of sunlight energy to electrical energy is determined by a Carnot heat engine. This limit is about 94%, based on the high temperature of the sun's black body at 5,800 K, and the cell temperature of about 350 K. There is, therefore, significant room to surpass the Shockley–Queisser limit.

Most of the solar cells manufactured today are made of crystalline silicon, and have shown efficiencies as high as 23% at a production-scale, although most cells are within the 17–19% range. In this Insight, Graetzel *et al.*[51] review how solution-processed solar cells could be competitive if the cost could be reduced and efficiency increased, but this is just one approach, competing with other rapidly advancing techniques.

Polycrystalline-silicon photovoltaics typically have efficiencies in the range of 14–16%, although recent progress indicates that refinements in bulk polysilicon production could yield efficiencies similar to that of single-crystal silicon for half of the ingot material[52]. The cost of photovoltaic-grade silicon forms half the cost of a solar module. In conventional silicon-cell production, about half the silicon is lost in sawing the bulk silicon into 150–200-μm-thick wafers. With modest light-trapping techniques[53], the wafers can be less than 50 μm thick. Other approaches to lowering costs include single-crystal epitaxial growth and lift-off, pursued by Solexel of Milpitas, California[54], and kerf-less manufacturing of wafers with large and aligned grains in which the thin wafers are directly made[55,56].

The other competitive approaches are based on direct-gap semiconductors such as cadmium telluride and copper-indium-gallium-selenide (CIGS), in which only a thin film (less than 1 μm) is needed to absorb solar radiation. Thin-film solar cells can be deposited on light and low-cost substrates, reducing their overall cost and weight, which is an advantage for transportation and installation. However, thin-film deposition on heterosubstrates introduces defects in these films, which currently limit production-scale efficiency of cadmium-telluride and CIGS thin-film solar cells to 12–14% — their practical upper limit is 29%. If the defects in cadmium-telluride thin films could be reduced in production, this would increase the carrier lifetime and increase the efficiency[57]. This reduction requires an understanding of the origins of defects, as well as ways to reduce and control them. The role of interfaces is a crucial aspect of solar cells and requires significant investment in research and development.

In addition to scientific understanding, engineered online monitoring and control systems may be needed to improve manufacturing quality control and as a means of continual improvement. Currently, panels made of cadmium-telluride thin films have the lowest cost (about $0.70 W^{-1}). Owing to the financial discount that is applied to any new solar technology that is expected to perform for at least 20–25 years, any photovoltaic technology developed has to have the technological 'head-room' to be competitive with the projected progress of more established technologies. An increase in efficiency from 17% to 20% would significantly reduce the panel cost and some of the balance-of-system costs, which could have a transformative effect on the market similar to that expected for polycrystalline silicon cells with an efficiency of over 18% (ref. 52).

Thin-film solar cells made of direct-gap III–V semiconductors are attractive because of their higher carrier mobilities, and their superior material quality that can make use of the materials research that focused on optoelectronics. This allows III–V semiconductor cells to achieve a much higher efficiency that is close to the practical upper limit. For example, solar-cell efficiencies of about 29% have been shown for III–V devices, which approaches the Shockley–Queisser limit of 33.5% for any single bandgap material[58]. Although high-quality films have generally been achieved on crystalline substrates, use of bulk materials would be cost-prohibitive for solar applications. The main challenge is to achieve the crystalline quality of III–V thin films, but on low-cost substrates such as plastics or sheet metals. As for techniques in solar cells that use silicon epitaxial growth followed by lift-off, good progress is being made towards manufacturing, on a larger scale, lift-off of thin crystalline gallium-arsenide layers grown epitaxially from bulk crystals.

Improving solar-module efficiency may also come from multijunction cells that capture a larger fraction of the solar spectrum, multi-exciton generation, multiphoton absorption or photon up-conversion, and light concentration. The use of quantum dots and molecules to utilize some of these phenomena are being pursued, and,

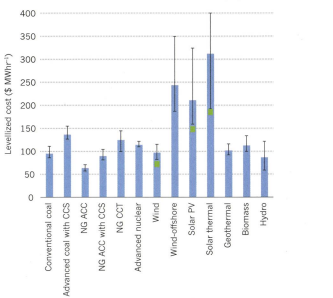


Figure 6 | Projections by the US Energy Information Administration made in 2010 of the 2016 average cost and range of the levellized cost of electricity from various sources. Note the cost of onshore wind at level 4 sites where the wind speed is 7.0–7.5 ms^{-1} and there are no financial incentives. Wind, utility-scale solar photovoltaics and solar thermal energy in 2012 in the United States (green) is already at or below the lower end of the 2016 estimates. ACC, advanced combined cycle; CCS, carbon capture and storage; CCT, conventional combustion turbine; NG, natural gas; PV, photovoltaics.

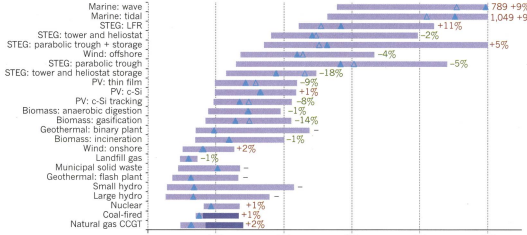


Figure 7 | Levellized cost of energy for quarter three, 2012. The analysis of cost in dollar per megawatt-hour is based on available data from projects in which investment or purchase contracts have been completed. To produce comparisons of levellized cost of energy at a global level, the analysis assumed a single corporate tax rate of 35%, an annual inflation of 2% and the equity investor internal rate of return of 10%. Quarter three central scenarios, filled triangles; quarter two central scenarios, open triangles. Adapted with permission from ref. 41.

as described by Graetzel *et al.*[51], are leading to a deeper understanding of light–matter interactions. The competition in solar technologies is fierce, and to have an effect on solar-electricity generation, the science must be translated into engineered systems that can meet the techno-economic target of about $0.50 W^{-1} for a solar panel at commercial scale.

Although the discussion so far has focused on solar flat panels, there are opportunities to generate electricity at $50–60 MWh^{-1} using concentrated solar photovoltaics. Most of the cost is in the balance of systems including the mirrors, lenses and their tracking system; however, the cost of the solar cell at the focal point of the optical system is not as significant. This is when multijunction solar cells with efficiencies that exceed 40% would be ideal. Innovations in low-cost and high-quality mirrors[59], inexpensive Fresnel optics, as well as creative engineering to track them throughout the day at low cost are being developed.

Complementary to electricity generation using photovoltaics, solar thermal approaches that concentrate sunlight in parabolic troughs and power towers have also been used to convert absorbed solar radiation into heat (typically below 450 °C). Electricity is then generated at about 30–35% efficiency using a heat engine[46]. Heat can also be stored to allow better demand matching, especially the offset peak electricity demand, by several hours in climates and economies in which air-conditioning demand is rising.

To reduce the LCOE to $50–60 MWh^{-1}, the efficiency of the heat engines must be increased. This requires more than 600 °C, suggesting that the higher-light concentration, as achieved in power towers or parabolic dish mirrors, is needed. There is significant opportunity for research to identify ways to store heat through the enthalpy of phase transitions or chemical reactions at more than 600 °C, as well as systems engineering to efficiently store and access this enthalpy on demand to run the heat engine for power generation[60]. Furthermore, research into low-cost and high-temperature materials, and their use in engineered systems, is needed to design and operate solar thermal reactors at these unprecedented high temperatures.

Although most of the work on concentrated solar power has involved either photovoltaic or thermal approaches, there are opportunities to combine them. For example, the visible part can be spectrally filtered out of the concentrated solar radiation for photovoltaics, and the rest of the radiation used for solar thermal approaches. Furthermore, solar thermal techniques rarely use all of the concentrated solar radiation, because there is generally some spillover from the solar absorber. Integrating photovoltaics in a solar thermal absorber may be possible, allowing both approaches to be integrated. Achieving more than 600 °C in a solar thermal absorber would also make it compatible with gas turbines used in natural-gas combined cycle technology, which would allow integration of solar and natural-gas power plants to address baseload and peak demands.

To convert direct current to alternating current electricity and to integrate solar power into the grid, power electronics is crucial and must be reduced to $0.10 W^{-1} or less — roughly half the cost it is today[46]. We are currently witnessing a revival of power-conversion electronics in which the introduction of wide bandgap semiconductors such as silicon-carbide and gallium-nitride, low-loss soft magnetics, solid-state capacitors and new-circuit topologies is having multiple positive effects[61]. These allow a significant increase in the operating frequency, which reduces the sizes of inductors and capacitors without altering their electrical impedance. The size reduction can allow integration of various components onto the same platform, reducing cost and weight. The use of solid-state capacitors as opposed to electrolytic ones can significantly increase reliability, which reduces the LCOE of solar energy.

The need to address the variable nature of intermittent technologies such as solar will become crucial at high levels of penetration. Approaches to forecast the time-dependent solar insolation are needed to plan electricity dispatch. In parallel, storage technologies that will offset the variations in solar electricity output need to be developed (see later).

Wind

The cost decline of wind energy is a result of significant improvements to turbines, blades and gearboxes, and an increase in the height of the wind towers. Although the largest wind turbine has a capacity of 7.5 MW, most turbines being installed are 1.5–2 MW (ref. 62).

As wind turbines increase in height and size, there is a marked increase in the mechanical stresses on the gearbox, blades and tower. Direct-drive wind turbines that can convert the slow rotary motion directly into alternating current electrical power using electromagnetic generators are receiving increased attention. However, these generators currently use permanent magnets based on rare-earth materials such as neodymium and dysprosium for their high remnant fields, field anisotropy and high-temperature stability. Although rare-earth deposits are found in many countries around the world, more than 90% of current production is in China[63,64].

Research in magnetic materials that can produce large magnetic energy density without the need for rare-earth metals is needed. For example, there are certain phases of iron-born compounds (for example, $Fe_{16}N_2$) that produce extremely high magnetocrystalline anisotropy, but that are difficult to access and stabilize under ambient conditions[65,66]. Today's computing power could be used to predict magnetic behaviour in a combinatorial search that could guide the discovery of magnetic materials with high crystalline anisotropy and those with accessible stable phases. Often, accessing these metastable phases requires nanostructuring, as well as manufacturing composites with materials such as soft magnets[67].

Offshore wind farms are an attractive source of energy because they can be placed in near-proximity to population centres and yet far enough away to mitigate local opposition. However, the cost remains high because of the need to build stable marine structures that can withstand the harsh marine environment for many years without the

need for replacement. Furthermore, because heavy-duty maintenance currently needs calm seas, offshore wind turbines have to be highly reliable. One approach to servicing the highest wind turbines is to design cranes that can 'inch worm' themselves up the towers, analogous to the self-lifting cranes used to build tall buildings.

Decarbonizing fossil-fuel emissions

If the world is to continue to generate electricity from fossil fuels, the carbon emissions from the major stationary 'point' sources will have to be significantly reduced in the coming decades. Achieving higher primary energy use through cost-effective improvements in thermodynamic efficiency, co-generation of power and heat, and use of waste heat can reduce carbon emissions. However, maximizing efficiency is not sufficient. Carbon capture and storage (CCS) from coal and natural-gas power plants and other major emitters, such as cement and steel plants, will be necessary.

A detailed analysis of post-combustion carbon capture science, engineering and economics can be found in a recent American Physical Society (APS) report[68]. The cost of CO_2 capture can be largely divided into two components: the capital cost, mostly of a sorbent-based CO_2 separation unit, the sorbent regenerator and the CO_2 compressor; and the operating cost, mainly a result of the parasitic energy needed to capture and compress CO_2. The capital cost depends on the size of the separation unit, which is inversely correlated with the rate constant of the CO_2 binding reaction with a sorbent. The energy cost, however, depends on the enthalpy of the sorbent-CO_2 dissociation reaction. The kinetics and the thermodynamics of the binding and unbinding reactions of today's sorbents are correlated (Fig. 8).

Research is needed to identify sorbents that have rate constants of more than 10,000 $M^{-1}s^{-1}$, but of enthalpies 40–80 kJ mol^{-1}. There are various secondary and tertiary amines, metal-organic frameworks and ionic liquids whose reactions could perhaps be tailored to approach desirable kinetic and thermodynamic performance levels[68] (Fig. 8). In biological organisms, enzymes such as carbonic anhydrase greatly reduce the free-energy barrier in the capture and release of CO_2. Researchers are making progress in genetically altering the properties of carbonic anhydrase[69] so that it can be used to improve the kinetics of low-enthalpy sorbents under the harsh conditions of the exhaust of a coal-fired power plant.

The cost of retrofitting existing pulverized coal plants to achieve 90% CO_2 capture with the technology available is estimated to require capital expenditures that approach those of the original plant[3]. In addition, 20–40% of the plant's energy would be diverted for separation, compression and transmission of the CO_2, so only the most thermodynamically efficient supercritical and ultra-supercritical plants could be considered for retrofits.

CO_2 capture can be achieved relatively easily if a highly concentrated stream of CO_2 exhaust is produced, which is possible by oxycombustion (in pulverized coal or natural-gas power plants) or integrated gasification combined-cycle (IGCC) technology. Cryogenic separation of oxygen from air is the most common industrial-scale method of production.

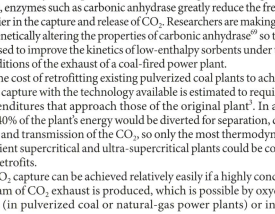


Figure 8 | Rate constant of current absorbents or adsorbents is about linearly proportional to the reaction enthalpy. New sorbents are needed to enhance the rate constant but maintain the energy required to release CO_2, to drive down the capital and operating costs of carbon capture. Adapted with permission from ref. 68.

There is no clear choice between oxycombustion and post-combustion capture[3], but it is clear that innovative chemical and physical separation methods need to be developed, piloted and demonstrated at near-commercial scale before wide-scale deployment can occur. Other economical solutions to CCS are needed, but without the prospect of a meaningful price on CO_2, the private sector is less motivated to invest in technology development. For this reason, research and development of CCUS (carbon capture, utilization and storage of carbon) in enhanced oil recovery and possibly enhanced geothermal energy[70] provides a financial incentive to invest in carbon-capture technologies.

Advancing CCS and CCUS could come from high-performance computing for *in silico* optimization of system performance, rather than expensive pre-commercial projects. High-fidelity simulations can also be used to reduce the risk of scaling small demonstration projects to larger systems, and thereby accelerate these projects adoption by industry.

Nuclear

Nuclear power can have an important role in efforts to decarbonize the production of electricity. Worldwide, nuclear energy constituted about 14% of the total electrical power generated in 2009 (ref. 71), but in 2011 this percentage dropped to about 12%, with a loss of 180 TWh of generation from Germany and Japan[3,72]. The reaction to the 2011 Fukushima nuclear disaster has been varied: some countries that have nuclear reactors are proceeding with caution, others continue to show interest in starting a nuclear-power programme, some countries have put their plans on hold and some have opted to reverse their decision on nuclear energy or accelerate planned phase-outs.

An increasing number of older nuclear plants will need to have their operating licenses extended beyond their original time limit if they are to continue operating. The cost of operating fully amortized plants in the United States is estimated to be about $23 MWh^{-1}, including the contribution of $1 MWh^{-1} to the nuclear-waste fund for the cost of spent fuel disposal[73]. The estimated cost of decommissioning each plant, including used fuel and site restoration costs is about $500 million per plant. The low cost of operating the current fleet of reactors has to be balanced with safety concerns, disposition of spent fuel and risks of nuclear weapons proliferation. Unfortunately, space does not allow a full discussion of these important concerns. Although there are significant challenges, the development of carbon-free sources of baseload power (nuclear and fossil plants with CCS) are needed to mitigate the worst risks of climate change.

Third-generation nuclear-power plants are engineered to be significantly safer than previous generation ones. In the event of an emergency shutdown of a nuclear reactor, the first few days — when the heat from daughter nuclear decay is largest and has to be dissipated — are the most critical. New reactors being constructed are designed to safely cool down for three days without access to grid or emergency electrical power or to water outside the reactor facility. The ultimate goal is to achieve passive safety, so there is very little chance of thermal damage to the reactor without requiring any active control after sudden shutdown.

The LCOE of new nuclear plants is about $100–120 MWh^{-1}; although this cost is not competitive with natural gas, it is less than estimates for fossil-fuel plants with CCS[4]. Another challenge is that the size of a single reactor is 1.0–1.5 GW with associated cooling and electrical distribution requirements. The cost of nuclear-reactor engineering, procurement and construction, operation and maintenance, and decommissioning is about $6,000–6,600 per installed kilowatt, which is roughly five times higher than comparable natural-gas generation[74] capital costs, and is quantized in increments of about $6 billion to $10 billion. The large financial exposure and the risk of construction and licensing delays have driven up the cost of financing plants[75].

The US DOE intends to support the engineering designs required by the Nuclear Regulatory Commission for licensing small-modular

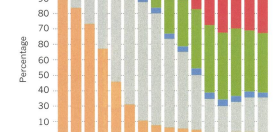


Figure 9 | History of the relative mix of the main source of energy used in the United States. The timescale for a shift in the energy mix is measured in decades. The growth in use of oil was increasingly driven by the transportation sector, for which it makes up 94% of transportation energy. Electricity is generated from coal (48%), natural gas (19%), nuclear power (21%), renewables 10%, and oil (1%). Adapted with permission from ref. 80.

reactors between 80 MW and 300 MW (refs. 76, 77). It is possible that safe nuclear power can be made more accessible through the economy of constructing dozens of reactors in a factory rather than one at a time at each site. Also, with the risk of licensing and construction delays reduced, small-modular reactors may represent a new paradigm in nuclear construction. The US DOE has also established an Energy Innovation Hub to develop multiphysics computational simulation tools to reduce the time needed to design and certify many aspects of both conventional reactors and small-modular reactors[78].

Grid energy storage

The reliability and efficiency of the transmission and distribution system is improved by grid energy storage. Balancing and optimizing power flow and generation are challenges that need improved technology, and business and regulatory vision. Significant improvements would help to realize more efficient energy markets, a lower cost of energy, increased reliability and security, and penetration of renewable energy beyond 20%.

The mismatch between an intermittent electricity supply and demand occurs over multiple time and energy scales. For seconds-to-minutes scales, voltage and frequency stabilization is currently achieved by grid-ancillary services that mostly use large banks of capacitors and hot-spinning reserves from fossil-fuel plants, but also banks of batteries and mechanical fly wheels. Hot-spinning reserves are also used to shore-up generation from minutes to hours. Modern natural-gas carbon-cycle plants are designed to ramp up rapidly at about 50 MW min^{-1}, which, when combined with improved sensing and control of the grid, will be sufficient to accommodate most of the intermittency problems of renewable energy. For hours-to-overnight scales, power and energy capacities of hundreds to thousands of megawatt and megawatt hour will be needed, alternatively many smaller (hundreds to thousands of kilowatts) of distributed energy storage sites can be used.

The most inexpensive method to store electricity for the medium- and long-term is to use pumped hydroelectric and compressed air in air-tight caverns, for which the capital cost is about $100 kWh^{-1}. However, geographical constraints warrant alternative approaches. New electrochemical and modular and low-cost fabricated compressed-air systems are promising[79].

Perspective

Given the anticipated technological advances in finding and extracting fossil fuels, the cost of these forms of energy could remain competitive with carbon-free sources for decades. In addition, the mix of energy sources in the United States (Fig. 9) indicates that the timescale of change will be roughly half a century[80]. If we are to mitigate climate risks in a timely manner, the speed of the adoption of clean-energy technologies at a widely-deployed commercial scale has to be accelerated. Although the deployment of any innovation will depend on its benefits relative to established technologies, inertia as well as real and perceived financial risks always favour the status quo. To accelerate the use of economically viable clean-energy solutions, government policies are needed to stimulate invention, innovation, and align the market forces.

With increases over the past 30 years in the number of extreme weather events, such as temperature extremes, floods, wildfires, droughts and storms, the overall loss trend is beginning to exceed $150 billion per year[81]. There is increasing evidence that these changes are linked with climate change[82]. Although the full costs of mitigating or adapting to these changes have significant uncertainties, prudent risk management does not equate uncertainty with inaction. Policies are needed that begin to include the full cost of the various sources of energy more directly into their market prices[82].

This Perspective highlights a few of the challenges and opportunities for science, engineering and policy. Our ability to find and extract fossil fuels continues to improve, and economically recoverable reservoirs around the world are likely to keep pace with the rising demand for decades. The Stone Age did not end because we ran out of stones; we transitioned to better solutions. The same opportunity lies before us with energy efficiency and clean energy.

The cost of renewable energy is rapidly becoming competitive with other sources of energy. With additional engagement of the scientific, financial and public-policy communities, as well as the general public, we can accelerate the transition to affordable, accessible and sustainable energy that will power economic growth, increase energy security and mitigate the risks of climate change. On the other hand, to quote from the IEA's World Energy Outlook, "If we don't change direction soon, we'll end up where we're heading". ■

1. Lee, R. The outlook for population growth. *Science* **333,** 569–573 (2011).
2. International Energy Agency in *World Energy Outlook 2011*, 546–547 http://www.worldenergyoutlook.org (International Energy Agency, 2011).
3. Committee on America's Energy Future, National Academy of Sciences, National Academy of Engineering and National Research Council. *America's Energy Future: Technology and Transformation* (National Academies Press, 2009).
4. US Department of Energy. *Report on the First Quadrennial Technology Review.* http://energy.gov/sites/prod/files/QTR_report.pdf (US DOE, 2011).
5. BP. *BP Statistical Review of World Energy.* http://bp.com/statisticalreview2012. (BP, 2012).
6. Maugeri, L. *Oil: The Next Revolution. Discussion Paper 2012–10.* http://belfercenter.ksg.harvard.edu/files/Oil-%20The%20Next%20Revolution.pdf (Belfer Center for Science and International Affairs, Harvard Kennedy School, 2012).
7. Inderwildi, O. & King, D. (eds). *Energy, Transport, & the Environment, Addressing the Sustainable Mobility Paradigm.* (Springer, 2012).
8. US Department of Energy. *Quadrennial Technology Review report, vol. II Technology Assessments* http://energy.gov/articles/department-energy-releases-inaugural-quadrennial-technology-review-report (US Department of Energy, 2011).
9. Gibbs, J., Pesaran, A. A., Sklad, P. S. & Marlino, L. D. in *Fundamentals of Materials for Energy and Environmental Sustainability* (eds Ginley, D. S. & Cahen D.) Ch. 31, 426–444 (Cambridge Univ. Press, 2012).
10. Powers, W. F., Automotive materials in the 21st century. *Adv. Mater. Process.* **157,** 38–41 (2000).
11. Cuddy, M. R. & Wipke, K. B. *Analysis of Fuel Economy Benefit of Drivetrain Hybridization.* http://www.nrel.gov/vehicleandfuels/vsa/pdfs/22309.pdf (National Renewable Energy Laboratory, 1997).
12. Holmberg, K., Andersonna, P. & Erdemirb, A. Global energy consumption due to friction in passenger cars. *Tribol. Int.* **47,** 221–234 (2012).
13. Schafer, A., Heywood, J. B. & Weiss, M. A. Future fuel cell and internal combustion engine automobile technologies: a 25-year life cycle and fleet impact assessment. *Energy* **31,** 2064–2087 (2006).
14. Manley, D. K., McIlroy, A. & Taatjes, C. A. Research needs for future internal combustion engines. *Physics Today* **61,** 47–52 (2008).
15. Yang, J. & Caillat, T. Thermoelectric materials for space and automotive power generation. *MRS Bull.* **31,** 224–229 (2006).
16. Scrosati, B. & Grache, J. Lithium batteries: status, prospects and future. *J. Power Sources* **195,** 2419–2430 (2010).
17. McCone, A. *Electric Vehicle Battery Prices Down 14% Year on Year* http://www.bnef.com/PressReleases/view/210 (Bloomberg New Energy Finance, 2012).
18. Chan, C. C. The state of the art of electric, hybrid, and fuel cell vehicles. *Proc. IEEE* **95,** 704–718 (2007).
19. Lukic, S. M., Cao, J., Bansal, R. C., Rodriguez, F. & Emadi, A. Energy storage systems for automotive applications. *IEEE Trans. Ind. Electron.* **55,** 2258–2267 (2008).
20. Spendelow, J. & Papageorgopoulos, D. Progress in PEMFC MEA component

R&D at the DOE fuel cell technologies program. *Fuel Cells* **11,** 775–786 (2011).
21. Debe, M. K. Electrocatalyst approaches and challenges for automotive fuel cells. *Nature* **486,** 43–51 (2012).
22. Yang, J., Sudik, A., Wolverton, C. & Siegel, D. J. High capacity hydrogen storage materials: attributes for automotive applications and techniques for materials discovery. *Chem. Soc. Rev.* **39,** 656–675 (2010).
23. Paster, M. D. *et al.* Hydrogen storage technology options for fuel cell vehicles: well-to-wheel costs, energy efficiencies, and greenhouse gas emissions. *Int. J. Hydrogen Energy* **36,** 14534–14551 (2011).
24. David, E. An overview of advanced materials for hydrogen storage. *J. Mater. Process. Technol.* **162–163,** 169–177 (2005).
25. Murray, L. J., Dincă, M. & Long, J. R. Hydrogen storage in metal–organic frameworks. *Chem. Soc. Rev.* **38,** 1294–1314 (2009).
26. Abbas, H. F. & Wan Daud, W. M. A. Hydrogen production by methane decomposition: A review. *Int. J. Hydrogen Energy* **35,** 1160–1190 (2010).
27. Krupnick, A. J. *Will Natural Gas Vehicles be in Our Future.* http://www.rff.org/RFF/Documents/RFF-IB-11-06.pdf (Resources for the Future, 2011).
28. NACS. *Key Facts and Figures.* http://www.nacsonline.com/NACS/Resources/campaigns/GasPrices_2011/Pages/default.aspx (NACS, 2011).
29. America's Energy Future Panel. *Liquid Transportation Fuels.* (National Academies Press, 2009).
30. Environmental Protection Agency. *Renewable Fuel Standard Program (RFS2): Final Rulemaking* http://www.epa.gov/otaq/fuels/renewablefuels/regulations.htm (US Environmental Protection Agency, 2010).
31. United Nations Environment Programme. *Towards Sustainable Production of and use of resources: Assessing Biofuels* http://www.unep.org/publications/contents/pub_details_search.asp?ID=4082 (United Nations Environment Programme, 2009).
32. Stratton, R. W., Wong, H. M. & Hileman, J. I. Quantifying variability in life cycle greenhouse gas inventories of alternative middle distillate transportation fuels. *Environ. Sci. Technol.* **45,** 4637–4644 (2011).
33. Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B. & Keasling, J. D. Progress in the microbial production of advanced biofuels: from feedstocks to fuels. *Nature* **488,** 320–328 (2012).
34. Worldwatch Institute. *Biofuels for Transport* (Routledge, 2007).
35. Georgianna, D. R. & Mayfield, S. P. Exploiting diversity and synthetic biology for the production algae biofuels, *Nature* **488,** 329–335 (2012).
36. Zhu, X-, G., Long, S. P. & Ort, D. R. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol.* **19,** 153–159 (2008).
37. Conrado, R. J. *et al.* in *Electrofuels: A new paradigm for renewable fuels. Advanced Biofuels and Bioproducts* (Lee, J. ed.) (Springer, 2012).
38. International Energy Agency. *World Energy Outlook 2011* 179,186 http://www.worldenergyoutlook.org/ (International Energy Agency, 2011).
39. International Energy Agency. *World Energy Outlook 2011*, 180 http://www.worldenergyoutlook.org/ (International Energy Agency, 2011).
40. US Energy Information Administration. *Annual Energy Outlook 2011 Levelized Cost Supplement* http://www.eia.gov/forecasts/archive/aeo11/electricity_generation.cfm (US Energy Information Administration, 2011).
41. Bloomberg New Energy Finance. *Levelised Cost of Energy Update, Q3 2012* http://www.bnef.com/WhitePapers/download/114 (Bloomberg New Energy Finance, 2012).
42. Wiser, R., Lantz, E., Bolinger, M. & Hand, M. *Recent Developments in the Levelized Cost of Energy from U.S. Wind Power Projects.* http://eetd.lbl.gov/ea/ems/reports/wind-energy-costs-2-2012.pdf (US Department of Energy, 2012).
43. Galen Barbose, G. Darghouth, N., Wiser, R. & Seel, J. *Tracking the Sun: An Historical Summary of the Installed Cost of Photovoltaics in the United States from 1998 to 2010* http://eetd.lbl.gov/ea/ems/reports/lbnl-5047e.pdf (Lawrence Berkeley National Laboratory report, 2011).
44. Solar Energy Industries Association. *US Solar Market Insight Report Q1 2012* http://www.seia.org/research-resources/us-solar-market-insight (Solar Energy Industries Association, 2012).
45. Hand, M. M. *et al. Renewable Electricity Futures Study* http://www.nrel.gov/analysis/re_futures/ (National Renewable Energy Laboratory, 2012).
46. US Department of Energy. *SunShot Vision Study* http://www1.eere.energy.gov/solar/pdfs/47927.pdf (US Department of Energy, 2012).
47. Bazilian, M. *et al. Re-considering the Economics of Photovoltaic Power* www.bnef.com/WhitePapers/download/82 (Bloomberg New Energy Finance, 2012).
48. Swanson, R. M. *Plenary talk at the DOE SunShot Grand Challenge: Summit and Technology Forum on June 14, 2012* https://www.eereenvisionsea.energy.gov/solar/videos/sunshot_summit_richard_swanson_plenary_session (US Department of Energy, 2012).
49. US Department of Energy. *SunShot Initiative* http://www1.eere.energy.gov/solar/sunshot/accelerating_deployment.html (US Department of Energy, 2012).
50. Swanson, R. M. *Proc. 31st IEEE Photovoltaic Specialists Conf.* 889–894 (2005).
51. Graetzel M. Jannsen, R. A., Mitzi, D. B. & Sargent, E. H. Materials interface engineering for solution-processed photovoltaics, *Nature* **488,** 304–312 (2012).
52. Stoddard, N., Wu B., Maisano, L., Russell, Clark, R. & Fernandez., J. *M. The Leading Edge of Silicon Casting Technology and BP Solar's Mono2 Wafers*, in *18th Workshop on Crystalline Silicon Solar Cells & Modules: Materials and Processes.* pp. 7–14. http://www.nrel.gov/docs/fy08osti/45745.pdf (National Renewable Energy Laboratory, 2008).
53. Mavrokefalos, A., Han, S. E., Yerci, S., Branham, M. S. & Chen, G. Efficient light trapping in inverted nanopyramid thin crystalline silicon membranes for solar cell applications. *Nano Lett.* **12,** 2792–2796 (2012).
54. Solexel. *Solexel unveils ultra-thin, high-performance silicon solar cell at Intersolar.* http://www.solexel.com/IntersolarNA2012.php (Solexel, 2012).
55. Rao, R. A. *et al. Proc. 37th IEEE Photovoltaic Specialists Conf.* 1504–1507 (2011).
56. Sachs, E. M., Wallace R. L., Hantsoo, E. T., Lorenz, A. M., Hudelson S.G.D. & Jonczyk, R. Methods for efficiently making thin semiconductor bodies from molten material for solar cells and the like, US Patent Application 20120067273 (2012).
57. Garland J. W., Biegala, T., Carmody, M., Gilmore, C. & Sivananthan, S. Next-generation multijunction solar cells: The promise of II-VI materials. *J. Appl. Phys.* **109,** 102423 (2011).
58. Miller, O. D., Yablonovitch, E. & Kurtz, S. R. Strong internal and external luminescence as solar cells approach the Shockley–Queisser limit. *IEEE J. Photovolt.* **2,** 303–311 (2012).
59. Angel, R. & Olbert, B. H. Method of manufacturing large dish reflectors for a solar concentrator apparatus. US Patent Application 20120125400 (2011).
60. Gur, I., Sawyer, K. & Prasher, R. Searching for a better thermal battery. *Science* **335,** 1454–1455 (2012).
61. Advanced Research Projects Agency-Energy. *ARPA-E Efficient Power Conversion Factsheet.* http://arpa-e-energy.gov/Portals/0/Documents/FundedProjects/ADEPT%20Slicksheets/ADEPT_Program_Final_One%20Pager.pdf (Advanced Research Projects Agency-Energy, 2010).
62. Bolinger, M. & Wiser, R. Understanding wind turbine price trends in the U.S. over the past decade. *Energy Pol.* **42,** 628–641 (2012).
63. US Department of Energy. *Critical Materials Strategy.* http://energy.gov/sites/prod/files/piprod/documents/cms_dec_17_full_web.pdf (US Department of Energy, 2010)
64. National Research Council. *Minerals, Critical Minerals, and the U.S. Economy* (National Academies Press, 2008).
65. Komuro, M., Kozono Y., Hanazono, M. & Yutaka, S. Epitaxial growth and magnetic properties of $Fe_{16}N_2$ films with high saturation magnetic flux density. *J. Appl. Phys.* **67,** 5126–5130 (1990).
66. Nian, J., Lawrence F. A., Edgar, T. C., &Wang, J.-P. N-site ordering effect on partially ordered $Fe_{16}N_2$. *Appl. Phys. Lett.* **98,** 092506 (2011).
67. Bader, S. J. Opportunities in nanomagnetism. *Rev. Mod. Phys.* **78,** 1–15 (2006).
68. American Physical Society. *Direct air capture of CO_2 with chemicals: a Technology Assessment for the APS Panel on Public Affairs.* www.aps.org/policy/reports/assessments/upload/dac2011.pdf (American Physical Society, 2011).
69. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nature Chem. Biol.* **5,** 567–573 (2009).
70. Pruess, K. Enhanced geothermal systems (EGS) using CO_2 as working fluid — a novel approach for generating renewable energy with simultaneous sequestration of carbon. *Geothermics* **35,** 351–367 (2006).
71. International Energy Agency. *Technology Roadmap — Nuclear Energy* http://www.iea.org/papers/2010/nuclear_roadmap.pdf (International Energy Agency, 2010).
72. Ruhl, C. *Energy in 2011 — disruption and continuity BP Statistical Review of World Energy* http://www.bp.com/sectiongenericarticle800.do?categoryId=9037130&contentId=7068669 (BP, 2012).
73. Nuclear Energy Institute. *Costs: Fuel, Operation and Waste Disposal* http://www.nei.org/resourcesandstats/nuclear_statistics/costs/ (National Energy Institute, 2012).
74. Nuclear Energy Institute. *The Cost of New Generating Capacity in Perspective* http://www.nei.org/resourcesandstats/documentlibrary/newplants/graphicsandcharts/the-cost-of-new-generating-capacity-in-perspective (Nuclear Energy Institute, 2012).
75. Massachusetts Institute of Technology. *Update of the MIT 2003 Future of Nuclear Power Study* http://web.mit.edu/nuclearpower/pdf/nuclearpower-update2009.pdf (Massachusetts Institute of Technology, 2009).
76. Chu, S. America's new nuclear option. *Wall Street J.* (March 23, 2010).
77. US Department of Energy. *Small Modular Reactor Fact Sheet* http://www.ne.doe.gov/pdfFiles/factSheets/2012_SMR_Factsheet_final.pdf (US Department of Energy, 2012).
78. US Department of Energy. *Consortium for Advanced Simulation of LWRs Project Summary.* http://www.ornl.gov/sci/nsed/docs/CASL_Project_Summary.pdf (US Department of Energy).
79. Advanced Research Projects Agency-Energy. *Grid-scale renewable energy storage factsheet.* http://arpa-e-energy.gov/Portals/0/Documents/FundedProjects/GRIDS%20Slicks/_GRIDS_ProgramSheet_Final.pdf (Advanced Research Projects Agency-Energy, 2010).
80. Energy Information Administration. *Annual Energy Review 2010.* Report No. DOE/EIA-0384(2010) www.eia.gov/totalenergy/data/annual/pdf/aer.pdf (Energy Information Administration, 2011).
81. Munich Re. *Topics Geo.* http://www.munichre.com/publications/302-07225_en.pdf (Munich Re, 2012).
82. Intergovernmental Panel on Climate Change. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation.* http://www.ipcc-wg2.gov/SREX/images/uploads/SREX-All_FINAL.pdf (Cambridge Univ. Press, 2012).

Acknowledgements The authors wish to acknowledge T. J. Augustine, P. Davis, H. Gruenspecht, M. Le, P. Lyons, R. Ramamoorthy, D. Sandalow, S. Satyapal and E. Toone for suggestions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/7pb7ff. Correspondence should be addressed to S.C. (the.secretary@hq.doe.gov) and A.M. (4majumdar.pubs@gmail.com).

REVIEW

doi:10.1038/nature11476

Materials interface engineering for solution-processed photovoltaics

Michael Graetzel[1], René A. J. Janssen[2], David B. Mitzi[3] & Edward H. Sargent[4]

Advances in solar photovoltaics are urgently needed to increase the performance and reduce the cost of harvesting solar power. Solution-processed photovoltaics are cost-effective to manufacture and offer the potential for physical flexibility. Rapid progress in their development has increased their solar-power conversion efficiencies. The nanometre (electron) and micrometre (photon) scale interfaces between the crystalline domains that make up solution-processed solar cells are crucial for efficient charge transport. These interfaces include large surface area junctions between photoelectron donors and acceptors, the intralayer grain boundaries within the absorber, and the interfaces between photoactive layers and the top and bottom contacts. Controlling the collection and minimizing the trapping of charge carriers at these boundaries is crucial to efficiency.

One hundred and twenty thousand terawatts of solar power irradiate Earth. Globally, humans consume only 15 terawatts. Solar power is abundant, free and can be harvested with minimal effect on the environment.

The cost and efficiency of the solar cells that make use of this power are of paramount importance. Single-crystal solar cells are formed from a continuous piece of crystal, but solution-processed photovoltaics are manufactured on large, potentially flexible substrates (Fig. 1). Light-absorbing and electrode materials are applied to the substrate by techniques such as blade coating, spray coating, printing and slot-die coating. Many solution-processing approaches use low temperatures, reducing the energy cost associated with solar-cell manufacturing and therefore the energy payback time. Lightweight, flexible solar modules can also lower the installation and maintenance costs.

Four main families of solution-processed solar technologies are approaching or have exceeded 10% solar-power conversion efficiency. The challenge now is to continue this progress. Advances rely on controlling the abundant materials interfaces that make up these highly nanostructured devices. In this Review, we discuss the chemistry, physics and materials science of these interfaces, and their control through materials chemistry. In particular, we focus on dye-sensitized solar cells (DSSCs)[1], organic photovoltaics[2], solution-processed bulk inorganic photovoltaics[3] and colloidal quantum-dot[4] solar cells.

The charge-separating junction is the boundary between two intimately contacted materials at which photogenerated electron–hole pairs are separated into longer-lived charge carriers (Fig. 2a, b). Efficient operation of the solar cell relies on the separation and collection of energy-bearing photocarriers before they lose their excess energy through recombination. Typical diffusion lengths for the performance-limiting photocarrier — minority charge carriers — are within the 5–500-nm range in solution-processed semiconductors, but the absorber thickness required to achieve complete absorption of solar irradiation of interest is often considerably more than this. This results in a possible trade-off between electron extraction and optical absorption if a planar absorber is used. To break this compromise, solution-processed solar cells make use of highly nanostructured and mesostructured interfaces.

Solution-processed photovoltaic materials generally contain significant degrees of disorder, often because of the boundaries that define the randomly orientated crystalline domains in both organic and inorganic semiconductors. Transport within these domains, as well as at their boundaries (including transport-limiting mechanisms, such as charge-carrier trapping), requires attention and optimization.

One example of a beneficial boundary effect is the quantum confinement of charge carriers to well-defined nanocrystalline domains, such as in semiconductor quantum dots. This effect allows the tuning of energy levels in materials. Wide-ranging tuning is advantageous because the Sun emits such a broad spectrum of light (Fig. 2c). Relying on one absorber bandgap limits the solar-cell power-conversion efficiency to 31% (ref. 5), whereas stacking cells with different bandgaps to create multijunction configurations could increase the power-conversion efficiencies to, in principle, a 68% limit (Fig. 2d).

We present the benefits of structured charge-separating interfaces and review the chemical management of their electronic properties, discuss interfaces within light-absorbing materials, describe the management and characterization of these interfaces, and summarize the progress and prospects in the field.

Morphology of charge-separating interfaces

DSSCs[6] (Fig. 3a) illustrate the use of nanostructured materials to overcome weak absorption. A monolayer of sensitizer molecules is anchored, with a linker, to an electron-accepting large-bandgap semiconductor such as titanium oxide, tin oxide or zinc oxide. If the monolayer was tethered to a planar electrode, it would absorb an insignificant fraction of solar radiation (less than 1%). Instead, because the electrode has a high-contact area, a three-dimensional junction is formed with the metal-oxide nanoparticles. This nanostructured electrode is many micrometres thick, and the dye–nanoparticle interface is more than a thousand times the area of the solar cell.

Interfacial considerations are of particular concern in an architecture in which every absorber molecule also forms one side of the charge-separating junction. Four interfacial electron-transfer reactions have a pivotal role in the efficiency of photoinduced charge separation. Injection of electrons from the dye into the conduction band of the oxide nanoparticles must proceed faster than exciton recombination within the dye. Recombination of forward-injected photoelectrons across the dye–electrode interface must be prevented. The sensitizer must be regenerated by donation of electrons from the electrolyte. Finally,

[1]Institute of Photonics and Interfaces, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland. [2]Molecular Materials and Nanosystems, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands. [3]IBM TJ Watson Research Center, Yorktown Heights, New York 10598, USA. [4]Department of Electrical and Computer Engineering, University of Toronto, Toronto, M5S 3G4, Canada.

recombination must be eliminated in the electrolyte or hole conductor. Through the molecular engineering of sensitizers, quantum yield values have exceeded 80% across the entire visible and near-infrared spectral range[7].

In organic solar cells, a mixture of semiconductors forms a bulk heterojunction[8,9] (Fig. 3b), again creating a large donor–acceptor interfacial area. The excitons generated through light absorption in the donor diffuse to a heterointerface with the acceptor, where they dissociate into weakly bound charge-transfer states or free carriers. These then need to be separated from the interface and transported to contacts through percolating paths within the distinct electron- and hole-transporting phases. This sequence of steps is affected by the morphology of the active layer. To optimize performance, intimate donor–acceptor interpenetration in the photoactive layers creates a large interface, resulting in a high quantum yield of exciton dissociation; these domains should be large enough to allow rapid charge separation and transport from the interface to prevent recombination[10]. Molecular ordering within materials and controlled phase segregation among the components enhance power-conversion efficiency.

The morphology that is needed to optimize performance is established during the formation of organic films. Solvent choice and drying, and processing additives, determine the nature and extent of the phase segregation. Post-deposition treatments such as thermal annealing may have a role in creating and stabilizing the nanoscale interpenetrating network of materials. For blends of poly(3-hexylthiophene) (P3HT) with the C_{60} fullerene derivative phenyl-C61-butyric acid methyl ester (PCBM) (Fig. 4a, b), both can be rendered semicrystalline. With annealing, P3HT forms long, thin conducting nanowires in a more homogeneous nanocrystalline PCBM film[11]. Increased phase separation and crystallization have a profound effect on the photophysical and transport properties of the active layer.

This insight has been gained through combined experimental and model-based analysis, such as electron tomography with optoelectronic modelling[12], and has led to the three-dimensional visualization of the bulk heterojunction, with implications for device performance (Fig. 4a). Bright-field transmission electron microscope (TEM) morphology observation and acquisition of a series of different tilt angles were combined and iteratively reconstructed to form three-dimensional TEM images, establishing the distance from each volume of donor material to the nearest location of acceptor material, known as the spherical contact distances. This determined the distribution of probability, assessed throughout the volume of a blended film, that an exciton generated inside the donor will find the acceptor at a certain distance. For excitons with a diffusion length of about 10 nm, the calculated exciton quenching efficiency was obtained for each sample, and the results showed that exciton quenching was much more efficient in materials with a finer bulk heterojunction morphology.

Colloidal quantum-dot solar cells (Fig. 3c) are also limited by the diffusion lengths of minority carriers (10–100 nm)[13,14], which are shorter than their above-bandgap optical absorption lengths (100–1,000 nm). They could also benefit from high interfacial areas. By moving beyond planar geometries, photocurrents have been increased because of the enhanced volume of depleted colloidal quantum-dot material from which electric-field-induced drift transport efficiently extracts photocharges[15]. In light of this, the optimal pore size is in the same range as the depletion-region thickness (150–200 nm). Because a continuous semiconductor phase of quantum dots rather than a monolayer of dye is used, the need for an abundant surface area is reduced compared with DSSCs, so columnar electrodes can be deployed[16].

Solution-processed copper-indium-gallium-selenide (CIGS) and copper-zinc-tin-sulphide (CZTS) semiconductors (Fig. 3d) form, with heat treatment, polycrystalline semiconductor films made up of large (micrometre-sized) single-crystal grains. As a result, electron and hole mobility can enter the 10–100 $cm^2 V^{-1} s^{-1}$ range, avoiding the need for bulk heterojunction analogues. In this case, planar interfaces are highly effective.

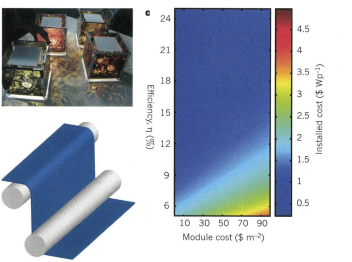


Figure 1 | Solution-processed photovoltaics. **a**, Commercial dye-sensitized solar cells illustrate the production of photovoltaic technologies through printing. During the day, the cells harvest light from the Sun or ambient light in a room, and in the evening the stored energy can be used for decorative illumination. Image courtesy of S. Uchida, University of Tokyo. **b**, Using a roll-to-roll technique to manufacture solution-processed solar cells reduces the manufacturing, deployment and energy costs involved. **c**, The installed cost per Watt-peak (Wp) as a function of module cost and module efficiency, assuming that balance-of-systems costs can be reduced to US$100 per square metre. Once the module cost is known and the efficiency determined, the installed cost can be established.

Electronics of charge-separating interfaces

When the interfacial area is maximized, the chemistry and resultant electronics of donor–acceptor interfaces become particularly important. On the positive side, the forward injection of charge carriers into the acceptor is helped by the abundance of sites available, but back recombination of the charge-separated carriers across the interface becomes more likely. Technical advances have highlighted the opportunities to manage the crucial charge-separating interface using materials chemistry.

In DSSCs, the light-absorbing dye molecules are anchored to the oxide nanoparticle electrode through coordinative bonding of the linker to the metal ions exposed at the surface of the metal oxide. Carboxylate and phosphonate groups are particularly effective as linkers. Forward transfer is determined by the energetic position of the lowest unoccupied molecular orbital (LUMO) level of the sensitizer compared with the conduction band of the acceptor, and the strength of electronic coupling between the initial and final states associated with the transfer. Engineering these two aspects has led to electron injection efficiencies approaching unity. Systematic investigations[17] of the dynamics of electron injection for perylene-carboxylic-acid-sensitized titanium-oxide nanocrystals have shown that forward injection proceeds with about a 13 fs time constant, which is consistent with theoretical predictions. Adding a saturated -CH2-CH2- spacer between the chromophore and the carboxylate anchoring group extended the injection time to 57 fs, confirming the strong distance dependence of the electron transfer rate. By contrast, introducing an unsaturated -CH=CH- did not alter the electron injection time, as a result of the enhanced electronic coupling through the double bond.

Recombination across the phase boundary that separates the electron- and hole-conducting media is the reverse of forward injection, and lowers the net quantum yield. Avoiding this requires engineering of the interface to impede the backflow of electrons across the junction. The sensitizer slows down the interfacial charge recombination by forming a dense self-assembled monolayer. This also occurs with infilling, using co-adsorbents at the oxide interface. This improved rectification benefits the open-circuit voltage and results in a higher net charge collection efficiency by ensuring that forward transfer is

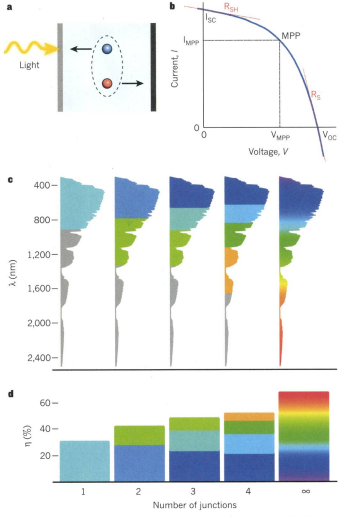


Figure 2 | Photovoltaic principles and performance. a, Light absorbed by the dye, polymer or inorganic semiconductor creates an excited photocarrier that quickly thermalizes to the band edge. Efficient solar cells extract the photocharge before it recombines. **b**, Solar cells must transfer power to load. Their operating current at the maximum power point multiplied by their operating voltage governs efficiency. The fill factor benefits from a minimal series resistance (R_S) and a maximum shunt resistance (R_{SH}). MPP, maximum power point; OC, open circuit; SC, short circuit. **c**, The solar spectrum is broadband, evenly dispersed across visible and infrared components. The optimal single-junction solar cell has a bandgap that corresponds to 950–1,250 nm. **d**, The theoretical efficiency of solar cells can be improved by sequentially harvesting the Sun's constituent spectral components in tandem, using multijunction cells that reduce losses associated with intraband relaxation.

quicker than recombination, achieving near-unity net collection of photogenerated electrons[18].

Impedance spectroscopy is an important tool for investigating electron transport and recombination in mesoscopic titanium-dioxide films, as well as interfacial charge transfer at the counter electrode and diffusion in the electrolyte. The shapes of the frequency-dependent impedance response curves allow clear identification of each of these processes. By using a transmission line model[19], electrical cell parameters, such as the resistances for electron diffusion and recombination, the chemical capacitance of the films, the diffusion coefficient of the redox mediator in the electrolyte, the resistance for electron transfer at the counter electrode and the series resistance of the cells can be extracted. Impedance parameters show the energetic distribution of the localized states in the bandgap of the oxide nanoparticles. Electrons move between the trap states and the conduction band, and only free electrons contribute to the steady-state diffusion process. The dependence of the recombination resistance on the bias potential follows an exponential function, suggesting a trap state-mediated charge-transfer mechanism. The width of the energy distribution of trap states can be quantified, and a prevalence of shallow traps, for a given open-circuit voltage, can increase the maximum extractable power from the cell, or fill factor[20]. Progress is still needed to quantify accurately the charge-carrier diffusion length[21], and to apply impedance spectroscopy to solid-state sensitized heterojunction solar cells[22] and colloidal quantum-dot sensitized solar cells[23].

Understanding and controlling the energetics of the charge-separating interface has contributed to the rapid improvement of colloidal quantum-dot solar cells. Early advances in performance relied on a metal-semiconductor Schottky junction, in which a light metal with a shallow work function established a charge-separating electric field within the p-type colloidal quantum-dot solid[24]. This model system was easily fabricated, rapidly achieved a multi-per-cent level of performance[25], and resulted in the rapid accrual of knowledge of doping and the establishment of charge-carrier drift and diffusion regions in colloidal quantum-dot solids[26].

Depleted heterojunction colloidal quantum-dot devices[27] (Fig. 3b) have a number of advantages over Schottky architecture. Separating charges on the light-absorbing side of the device lessened the reliance on long-distance transport within the quantum-dot film. The n-type titanium-oxide to p-type colloidal quantum-dot film heterojunction removed the limits on open-circuit voltage that arose as a result of Fermi-level pinning in Schottky devices. The barrier-to-hole extraction presented by the large-bandgap titanium dioxide was used to prevent back recombination.

The electron affinity of the acceptor that makes up the depleted heterojunction has also been systematically varied[28]. Cyclic voltammetry studies of electrode and quantum-dot materials showed that only a mildly favourable band offset was required to allow efficient charge injection into the titanium dioxide. This is consistent with the small binding energy of electron–hole pairs in metal chalcogenides, and with the p–n junction also driving carrier extraction along with the heterostructure band offset[29]. Multilayer electrodes in colloidal quantum-dot solar cells have been exploited to allow greater freedom in the joint optimization of doping and band offset[30].

Inorganic solution-processed solar cells, such as those based on CIGS and CZTS, are also reliant on the front junction for efficient charge separation. Here an n-type semiconductor, such as cadmium sulphide, forms a depletion region when in contact with the p-type absorber. A large positive conduction band offset (spike) (Fig. 3d) presents a barrier for minority carrier (electron) collection, reducing short-circuit current. By contrast, negative offset (cliff) leads to increased recombination at the buffer–absorber interface, thereby reducing open-circuit voltage. Ideally, the device would have a small 0–0.4 eV conduction-band offset spike, as found in copper-indium-gallium-sulphur-selenide (CIGSSe) devices employing a cadmium-sulphide buffer[31,32]. Efforts to replace cadmium sulphide with a buffer that contains no heavy metals must take into account the effect of conduction-band offset on device performance[33].

At the back of the CIGS or CZTS devices, an ohmic contact is required to collect majority carriers from the absorber. Molybdenum provides an ohmic contact for both CIGSSe and copper-zinc-tin-sulphur-selenium (CZTSSe) devices, as a result of the molybdenum-selenide interfacial layer that forms between the two, during high-temperature thermal processing[34]. Not only does the molybdenum act as a back contact, but it also provides a partial chemical barrier. Trace amounts of sodium significantly improve CIGSSe device performance[35], particularly open-circuit voltage, and this element is often introduced by fabricating devices on soda-lime glass. The flow of sodium into the absorber can be controlled during device fabrication by choosing a particular molybdenum-film thickness, density or grain structure, as well as through the application of a diffusion barrier to the glass.

In organic solar cells, the charge-transfer state is formed during rapid (about 40 fs)[36] exciton dissociation[37,38] at organic semiconductor

heterojunctions. The energy of the charge-transfer state has an important role in determining the open-circuit voltage of the cell. The rapidly ensuing dissociation into free charge carriers, which must outpace recombination, maximizes the short-circuit current. Maximizing open-circuit voltage requires organic or polymer heterojunctions in which the energy of the lowest intramolecular singlet excitons in the donor and acceptor is just sufficient to create the charge-transfer state, placing the system at the threshold for photovoltaic operation: synthetically this requires careful tuning of the highest occupied molecular orbital (HOMO) and LUMO levels to maximize the open-circuit voltage while still achieving efficient charge separation. Detailed photophysical studies have quantified the considerable difference in binding energy of charge-transfer excitons compared with bulk-charge separation[39], and indicate the importance of work aimed at directing charge flow away from the heterojunction to avoid losses as a result of bound charge-transfer excitons.

The electronic properties of the donor–acceptor interface in organic bulk heterojunction solar cells have been analysed using a combination of highly sensitive photothermal deflection spectroscopy, external quantum efficiency measurements and electroluminescence spectroscopy to obtain detailed information on the energy and role of the interfacial charge-transfer state in generating free charges[38,40]. Scanning tunnelling microscopy and spectroscopy probed the local photovoltaic performance of prototypical polymer–fullerene bulk heterojunction films with 10-nm resolution. Fullerene-rich clusters act as sinks, extracting electrons from an interface layer of a homogeneously mixed polymer–fullerene matrix, surrounding the fullerene cluster and allowing quantitative correlation between nanoscopic and macroscopic device photovoltaic performance[41].

As well as the donor–acceptor junction, the interface of the photoactive layer with the electrodes is important for the operation of organic solar cells. Once the HOMO and LUMO levels have been optimized inside the device, the difference in the work function of the electrodes can limit the maximum attainable open-circuit voltage. By incorporating thin interfacial layers of either alcohol- or water-soluble polymers[42] or a fullerene bisadduct surfactant[43] between the photoactive layer and the electron-collecting electrode, performance was substantially increased. These improvements have been attributed to the formation of dipole layers and to Fermi-level pinning, respectively. At the hole-collecting electrode, replacement of the standard electrode material, poly-3,4-ethylenedioxy-thiophene–poly(styrene sulphonate) (PEDOT–PSS), with nickel oxide (ref. 44) or a single layer of electronically tuned graphene oxide[45], has also resulted in significant improvements. Further research is crucial to develop solution-processed tandem and multijunction polymer bulk-heterojunction solar cells in which the individual subcells are stacked using optically transparent electrically conducting recombination layers[46,47].

Interfaces within the absorber

Much of the progress in colloidal quantum-dot photovoltaics has occurred within the quantum-dot solid itself, which has a large interfacial area. Early work in quantum-dot photovoltaics[48] assumed that a two-phase solution, similar to the bulk heterojunction of organic photovoltaics, could extend the excited state lifetimes. However, completely removing the polymer phase and organic insulating ligands led to marked improvements in transport performance[49], ensuring the focus was on essentially pure colloidal quantum-dot films. Only the shortest organic ligands were retained for surface passivation. Solution-phase exchange to short organic linkers — such as butylamine — took colloids to the edge of stability in non-polar solvents, and allowed sufficient communication among the nanoparticles to achieve efficient extraction over 100-nm length scales[25]. Solid-state exchanges were devised[50] in which a quantum-dot film is cast, and a crosslinking agent is introduced into a solvent that does not redisperse the solid-state quantum-dot film, but instead aids in the displacement and removal of the original ligands[24]. The organic molecule chosen plays an important

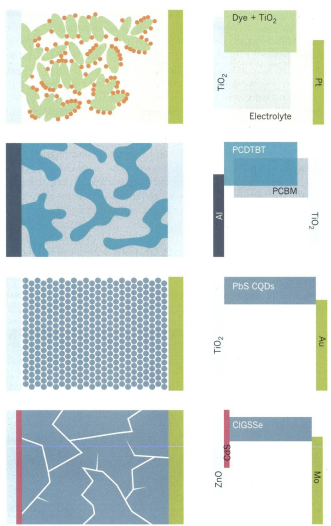


Figure 3 | Solution-processed solar cell architectures. **a**, In the dye-sensitized solar cell, light-absorbing dye (orange circles) is tethered to an electron-accepting three-dimensional electrode. The dye injects an excited photoelectron into the electrode and is recharged through an infiltrated electrolyte. **b**, The excitons generated in the light absorber of organic solar cells that benefit from an interpenetrating bulk heterojunction, rapidly diffuse to an interface where they are separated. This allows charge extraction prior to recombination. **c**, The light-absorbing region of colloidal quantum-dot solar cells consists of a densely-packed, well-passivated collection of monodispersed semiconductor nanoparticles that have a size-tuned bandgap. The quantum-dot layer transfers electrons to the titanium-dioxide electrode and transports holes to the back ohmic contact. **d**, The minority carriers in inorganic bulk polycrystalline solar cells diffuse through a polycrystalline semiconductor to be extracted at a heterojunction. The panels on the left illustrate the three-dimensional physical structure of the cells, the panels on the right show the flat band (open circuit) energetic structure of the cells.

part in establishing the film morphology, and the thiol-end functional group has been particularly effective in terminating nanoparticle surface dangling bonds[51]. The conductivity of the linker molecule itself may be important in enhancing transport in colloidal quantum-dot solids[13].

Because of the intense focus on inorganic semiconductor nanoparticles capped using organic ligands, enhanced passivation and retained quantum confinement using an inorganic passivation strategy came as a surprise[52]. Time-resolved transient infrared spectroscopy (Fig. 4e) indicated there was a crucial role for trap states, as opposed to the quantum-confined core levels, in determining the effectiveness of transport in colloidal quantum-dot films[52,53]. Halide anions introduced during solid-state film treatments have led to a marked reduction in the density of trap states deep within the bandgap of the colloidal quantum-dot solid.

This insight has been extended to the organic passivation domain, in which the mechanisms that underlie improved performance associated with optimal ligand selections have been traced directly to the quality of passivation in the final quantum-dot solid[23]. Optical field-effect transistor studies[54] of colloidal quantum-dot films have also suggested a midgap band. The work connects the relationship between dark transport and photoconductivity in films of lead-sulphide nanocrystals. Spectrally resolved photoresponses of these devices show that a weakly conductive midgap band is responsible for charge transport in the dark. When illuminated, the material becomes dominated by band-edge quantized states. The midgap band is important even under illumination because its occupancy controls the dynamics of band-edge charges. The colloidal quantum-dot field will benefit from studies that help to find, and explain, the relationships between electron and hole mobilities measured using field-effect transistors, time of flight, and charge extraction, by linearly increasing voltage techniques.

Solution-processed CIGSSe or CZTSSe devices also provides examples of chemical and interfacial effects within the semiconductor itself. Processing of thin-film chalcogenide compounds from ink-like solutions using techniques such as spin coating, slit casting and spray coating can control the metal stoichiometry, which is crucial for high device performance[55]. In CIGSSe devices, the indium to gallium ratio is used to tailor the bandgap, and the materials should be copper-deficient relative to the ideal chalcopyrite stoichiometry to avoid detrimental electrical shunting. Similarly in CZTSSe devices, films should be both copper-poor and zinc-rich to form the highest-performing devices[56]. Using solution-based approaches, the final metal composition can be determined from the initial quantities of the elements in the solution; although for CZTSSe devices, tin volatility and loss during thermal processing must be mitigated to achieve the highest performance[57]. For CIGSSe, the ease with which targeted impurities can be introduced into the deposition solution led to the discovery that low levels of antimony markedly improved grain structure and CIGSSe device performance[58] (Fig. 4c).

High performance relies on diffusion during CIGSSe or CZTSSe device processing, and homogeneity in the final film. In addition to sodium diffusing from the glass substrate[35], cadmium diffuses into the absorber to a depth of about 10 nm during cadmium-sulphide chemical bath deposition (studied mostly in CIGSSe devices), leading to n-type doping (Cd_{Cu} donor defects) that can invert the top layer of the absorber and increase the band bending, and hence electric field, near the semiconductor junction[59]. Deeper diffusion of cadmium has been reported in films with many grain boundaries — the internal interfaces within the absorber film[60].

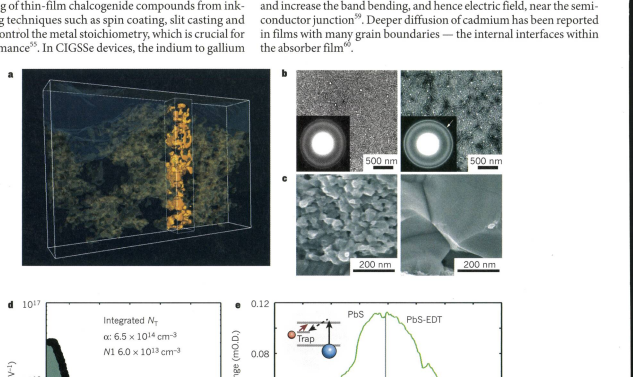


Figure 4 | Analysis of interface morphology and electronic properties. **a**, Three-dimensional electron tomography maps the detailed structure of bulk heterojunctions and guides optoelectronic modelling (Reprinted with permission from ref. 12). **b**, In the formation of a phenyl-C61-butyric acid methyl ester–poly(3-hexylthiophene) (PCBM) heterojunction, annealing conditions (right panel) control the length scale of interpenetration of the donor and acceptor materials (Reprinted with permission from ref. 11). **c**, Scanning electron microscope images of copper-indium-gallium-selenide (CIGS) semiconductors prepared with (right panel) or without a small concentration of antimony (Sb). Antimony improved the grain structure achieved in film processing (Reprinted with permission from ref. 58). **d**, Admittance spectroscopy shows the trap states in solution-processed CIGS devices. By correlating with materials analysis, materials chemistry mechanisms that underlie recombination centre formation could be pinpointed (Reprinted with permission from ref. 64). N_T, defect density of states; α, trap levels with activation energies at around 25 meV (green area); $N1$, trap levels with activation energies at around 110 meV (purple area). **e**, Colloidal quantum-dot solar cells have, until recently, relied mainly on organic molecule-based control over surface passivation. A highly compact halogen-based inorganic passivation strategy superseded the performance of the organic approach. Time-resolved infrared spectroscopy was used to explore how organic and inorganic ligands used in the passivation of colloidal quantum-dot solids affect electronic trap states (Reprinted with permission from ref. 53). EDT, ethane dithiol; IR, infrared; MPA, 3-mercaptopropionic acid; PbS, lead sulphide.

Grain boundaries are typically copper-poor in CIGSSe devices[61] relative to the grain interiors, which modifies the valence band offset such that the boundaries repel holes and recombination is reduced. Low-temperature air-annealing after device fabrication improves the performance of CIGSSe and CZTSSe devices, particularly for those with small grain size, as a result of grain-boundary passivation[62]. Finally, in the bulk absorber at the atomic scale, elemental vacancy and substitutional defects, such as V_{Cu}, In_{Cu} and Cu_{Zn}, significantly affect doping and recombination, and consequently device performance[63,64].

Trap spectroscopies have been used in solution-processed inorganic solar cells and have been combined with a suite of advanced-materials analysis methods[64,65]. The defect level distribution for solution-processed CIGSSe films grown with low selenium supply showed trap levels with activation energies of 25 meV and 110 meV (ref. 66) using admittance spectroscopy (Fig. 4d). Materials analysis has shown the importance of control over the supply of selenium during CIGSSe grain growth. The difference in selenium supply has a negligible effect on CIGSSe grain structures, crystal orientation and bulk composition, when measured by cross-section scanning electron microscopy or TEM, X-ray diffraction and secondary ion mass spectrometry. But when selenium supply is insufficient, a defective chalcopyrite layer with a high density of $(In,Ga)_{Cu}$ defects forms near the CIGSSe surface owing to the loss of selenium from the CIGSSe lattice. This defective chalcopyrite layer leads to more pronounced rollover and crossover signatures in J–V curves. The occupation of existing V_{Cu} by extra (In, Ga) atoms to form $(In, Ga)_{Cu}$ defects is also manifest in their defect spectra, in which a significant decrease in shallow acceptor defects (attributed to V_{Cu} defects) but an increase in the deeper $N1$ defect level (110 meV from the band edge) (attributed to $(In, Ga)_{Cu}$ defects) is observed. These defects affect solar cell performance by acting as Shockley–Read–Hall recombination centres. This insight can be used to further optimize CIGSSe processing, highlighting the importance of comprehensive electronic and materials characterization.

Spectral use to improve performance

In DSSCs, using co-sensitizers with black ruthenium dye has resulted in a power-conversion efficiency of 11.4% (ref. 67). A further increase in power-conversion efficiency to 12.3% has been reported[18]. This was achieved by using a judiciously engineered porphyrin sensitizer along with a cosensitizer and Tris(2,2′-bipyridyl)cobalt(III) as a redox shuttle[68] (Fig. 5). This redox shuttle yields substantially higher open-circuit voltage values than the iodide-based electrolyte. If validated by a photovoltaics calibration laboratory, the power-conversion efficiency would be the highest achieved for a DSSC under standard reporting conditions.

In organic photovoltaics, progress on small-bandgap polymers and small molecules has been chiefly responsible for achieving 8.4% efficiency in a single-junction low-bandgap polymer cell[42] and 8.6% in a tandem configuration[69]. Achieving these performance levels required synthesizing a small-energy-bandgap polymer (less than 1.5 eV) with carefully tuned HOMO and LUMO levels to maximize the open-circuit voltage, but still achieve efficient charge separation. Optimization of both charge-carrier mobility and fine-phase separation with the acceptor led to a high current density and fill factor. As yet undisclosed organic semiconductors have achieved record certified solar-cell efficiencies of 10.0% for single junctions[70] and 10.6% for tandem devices.

Efforts to improve colloidal quantum-dot solar photovoltaics are based on improved spectral use. The optimal single-junction bandgap close to a wavelength of roughly 1 μm is achieved using readily synthesized lead-sulphide quantum dots with about a 3-nm diameter. These devices have progressed rapidly from concept[48] to all-quantum-dot (polymer-free) devices based on Schottky junctions[13,25,71]. Depleted-heterojunction architectures[27] have achieved power-conversion efficiencies of 6% (ref. 52). The quantum-dot tandem cell is the first tandem to use the same material in both the large-bandgap front cell and small-bandgap back cell, with quantum-size-effect tuning alone allowing a wide variation in bandgaps (1.6 eV and 1.0 eV)[72]. This result relied on an interconnection strategy known as the graded recombination layer, which employed a stepwise progression in work functions from the shallow-work-function electron acceptor of the back cell to the deep-work-function hole acceptor of the front cell.

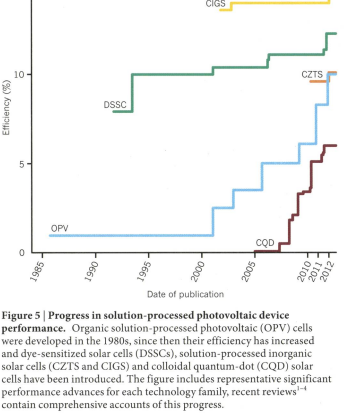


Figure 5 | Progress in solution-processed photovoltaic device performance. Organic solution-processed photovoltaic (OPV) cells were developed in the 1980s, since then their efficiency has increased and dye-sensitized solar cells (DSSCs), solution-processed inorganic solar cells (CZTS and CIGS) and colloidal quantum-dot (CQD) solar cells have been introduced. The figure includes representative significant performance advances for each technology family, recent reviews[1–4] contain comprehensive accounts of this progress.

CZTSSe and CIGSSe films control the bandgap by optimizing the sulphur to selenium ratio and metal substitution[73,74], which can further enhance spectral use and minimize recombination in this materials class.

Controlling heterointerfaces

The first organic photovoltaics devices were planar[75], but they soon progressed with the advent of the bicontinuous network of internal donor–acceptor heterojunctions[8] founded on rapid photoinduced electron transfer from the excited state of a semiconducting polymer onto C_{60} buckminsterfullerene. Refining the control over molecular morphology of the active layer[76,77] by improving process conditions had a principal role in increasing the power-conversion efficiency from a few per cent a decade ago[9] to more than 10% in 2012 (ref. 70). In colloidal quantum dots, both nanoparticle[15] and patterned-substrate electrodes[16] enhanced short-circuit currents by about 20%. In both cases, improved control over the aspect ratio (morphology), as well as transport and trapping, at these interfaces could provide further benefits.

Controlling the bulk semiconductor solid

Solution-based approaches provide good control over film stoichiometry. One of the first approaches for CIGSSe devices involved depositing metal-oxide nanoparticle-based precursors onto a substrate, chemically reducing the precursor film to the metals using an anneal in hydrogen, and subsequently selenizing or sulphurizing the film using a post-deposition treatment in a sulphur or selenium atmosphere[78], leading to CIGSSe devices with 13.6% efficiency. An analogous approach using chalcogenide-based nanoparticles subsequently improved efficiency to 14% (ref. 79). A pure-solution approach using hydrazine as the solvent led to 15.2% efficiency[80] for a fully solution-processed device, closing the performance gap with vacuum-deposited devices, which have a

20.3% efficiency. A similar hydrazine-based slurry approach for the CZTSSe system led to 10.1% efficiency, which is the highest achieved by any solution- or vacuum-based approach for this materials system, demonstrating the potential of solution-processed devices to compete with vacuum-based analogues in performance[81]. Analogous progress using a pure-nanoparticle approach — not requiring hydrazine — has led to an 8% efficient cell[24]. Stoichiometry and its control are essential to CZTSSe, and depositing material is complicated by the narrow range of phase stability, and its susceptibility to the loss of tin, sulphur and selenium during heat treatment[57].

The success of the hydrazine-based approach for CIGSSe and CZTSSe devices is as a result of the nanoscale mixing of the various constituents in the precursor film. Metal species already have the desired metal-chalcogen bonding, and incorporating excess sulphur or selenium in the film before heat treatment allows the rapid formation and crystallization of the CIGSSe or CZTSSe phase. Clean decomposition and dissociation of hydrazine from the precursor film during heat treatment also contributes to the formation of an impurity-free (especially oxygen and carbon) final film.

Outlook

Solution-processed solar cells are a promising technology for reaching 15% solar-power-conversion efficiency and beyond in low-cost, flexible solar modules. Inorganic solution-processed bulk semiconductor approaches have a tried-and-tested solar-materials composition achieved not through more costly physical deposition methods, but instead through solution-phase processing. Reducing the cost and complexity of solvents, their handling during processing and the temperature required in thermal processing are all being investigated. Organic solar cells and DSSCs have efficiencies in excess of 10%, and each has continued to progress in performance and reliability; for example, DSSC modules with 8% efficiency have passed tests for the outdoor use of photovoltaic panels, including long-term exposure to heat and humidity, light soaking and temperature cycling[82]. Life-cycle analysis suggests that the energy payback for a DSSC is less than one year in a southern European climate, compared with more than three years for silicon solar cells. The youngest of the technologies reviewed, colloidal quantum-dot solar cells, raises the possible performance further in view of their intrinsic compatibility with systematically engineered multijunction architectures, and the possible opportunity to exploit multi-exciton generation[83]. These materials require further improvements in transport and passivation to approach their full potential.

Improvements in efficiency will include addressing the open-circuit voltage deficit in solution-processed solar technologies, especially CZTSSe devices. The voltage provided by many solution-processed devices is appreciably lower than the theoretical maximum. In CZTSSe devices, the voltage seems to be limited by interface-dominated recombination[84]. Understanding the origins of this process requires a deeper understanding of the materials interfaces. In organic photovoltaics, at present, the open-circuit voltage is over 0.6 eV smaller than the bandgap[85], but in principle a difference as little as 0.35 eV is possible. Reducing the energy losses associated with exciton dissociation, charge separation and collection will be essential.

Reducing ohmic losses must be achieved in many solution-processed device classes. A higher series resistance in CZTSSe devices than in CIGSSe devices (especially for high-sulphur-content absorber layers) and an increase in series resistance with reduced temperature, suggests the presence of an energetic barrier to charge-carrier extraction, or that CZTSSe devices have a deeper acceptor than CIGSSe devices[62], leading to carrier freeze-out at low temperatures. If a barrier to current transport is present, it must be located at the back contact, the charge-separating interface or at interfaces of the polycrystalline grains that make up the solid, and remedied.

In colloidal quantum-dot photovoltaics, work is needed to enhance passivation within the bandgap of the quantum-dot solid, in which midgap and shallow traps of considerable energetic depth below the quantum-confined band edges are present. Modelling indicates that trap-free quantum dots are, in principle, possible[86]. Shallow traps originate because of surface disorder and reconstructions, whereas deep traps form because of significantly undercoordinated atoms on the surface[87–89]. These effects are analogous to the exponential decay that can occur near the band edge, known as Urbach tails, owing to deviations of bond lengths and bond angles, from ideal bulk values and unsaturated dangling bonds, respectively. Satisfying the charge-neutral conditions when selecting the intended stoichiometry of the quantum dots — similar to electron counting rules that describe the reconstruction of flat surfaces — offers a maximum opening of the quantum-dot bandgap[86]. Surface self-healing to remove dangling bonds reduces the total energy of the dots further[87]. These findings suggest that the dots have an intrinsic propensity to be trap-free in thermodynamic equilibrium conditions. In contrast, experimental work has used synthesis and ligand-exchange conditions that are highly non-equilibrium and kinetically driven, requiring a renewed effort in thermodynamically driven synthesis and exchange.

Progress in the control of structured interfaces is ongoing for photovoltaics that can benefit from high-surface-area charge-separating junctions. Mesostructures for enhanced light-harvesting are a particularly fertile opportunity, including nanorods composed of oxides with a core shell structure, gyroids and nanochannel-based mesoporous semiconductor oxide beads. These will facilitate electron transport and collection, and increase multiple light scattering, leading to an augmented optical path for a given amount of the light-absorbing moiety.

Improving the control of structured interfaces allows the introduction of light management techniques. In addition to random surface texturing of the electrode-absorber interfaces used in conventional solar cells to enhance absorption through multiple-angle scattering[90], photonic and plasmonic mesostructures can be incorporated into thin-film devices to increase the effective optical path lengths[91]. Examples are sub-wavelength metal nanoparticles, grating patterned metal films, semiconductor nanowires and photonic crystals[92]. Photocurrent enhancements of 33% have been demonstrated for thin-film plasmonic silicon devices[93]. Initial studies in plasmonic organic cells[94] and DSSCs[95] have resulted in efficiency enhancements of 20% and 50–100%, respectively, which are promising for light trapping in solution-processed technologies.

One opportunity for the field of solution-processed photovoltaics is to combine concepts and materials from diverse materials communities. For example, quantum dots could be used as replacements for chromophores in DSSCs, with a monolayer of size-tuned nanoparticles forming the broad spectral reach of semiconductors, and complemented by the well-understood charge transfer and back recombination blocking of the DSSC architecture. With their excellent efficiency in harvesting the visible portion of the Sun's spectrum, organic photovoltaics could be integrated with infrared-bandgap semiconductors[96], such as colloidal quantum-dot solids, as the back cell in tandem or multijunction designs. While these various opportunities and challenges are addressed, solution-processed photovoltaics will continue to help drive higher performance at lower cost for the widespread deployment of solar-harvesting technology. ■

1. Nazeeruddin, M. K., Baranoff, E. & Grätzel, M. Dye-sensitized solar cells: a brief overview. *Solar Energy* **85,** 1172–1178 (2011).
This article provides a review of DSSCs.
2. Nelson, J. Polymer: fullerene bulk heterojunction solar cells. *Mater. Today* **14,** 462–470 (2011).
This review is an overview of organic and polymer solar cells.
3. Todorov, T. & Mitzi, D. B. Direct liquid coating of chalcopyrite light absorbing layers for photovoltaic devices. *Eur. J. Inorg. Chem.* **2010,** 17–28 (2010).
This article reviews inorganic solution-processed solar cells.
4. Sargent, E. H. Colloidal quantum dot solar cells. *Nature Photon* **6,** 133–135 (2012).
This article is a recent review of colloidal quantum-dot solar cells.
5. Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p–n junction solar cells. *J. Appl. Phys.* **32,** 510–519 (1961).
6. O'Regan, B. & Grätzel, M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films. *Nature* **353,** 737–740 (1991).

This is a seminal paper on DSSCs, reporting an early major advance in performance.

7. Grätzel, M. Recent advances in sensitized mesoscopic solar cells. *Acc. Chem. Res.* **42,** 1788–1798 (2009).
8. Shaheen, S. E. *et al.* 2.5% efficient organic plastic solar cells. *Appl. Phys. Lett.* **78,** 841843 (2001).
This is the first report of a bulk heterojunction used in organic polymer solar cells.
9. Yu, G., Gao, J., Hummelen, J. C., Wudl, F. & Heeger, A. J. Polymer photovoltaic cells: enhanced efficiencies *via* a network of internal donor-acceptor heterojunctions. *Science* **270,** 1789–1791 (1995).
This is a seminal paper on polymer photovoltaics, reporting the benefits to be achieved at the bulk heterojunction.
10. Maturová, K., van Bavel, S. S., Wienk, M. M., Janssen, R. A. J. & Kemerink, M. Description of the morphology dependent charge transport and performance of polymer:fullerene bulk heterojunction solar cells. *Adv. Funct. Mater.* **21,** 261–269 (2011).
11. Yang, X. *et al.* Nanoscale morphology of high-performance polymer solar cells. *Nano Lett.* **5,** 579–583 (2005).
12. Oosterhout, S. D. *et al.* The effect of three-dimensional morphology on the efficiency of hybrid polymer solar cells. *Nature Mater.* **8,** 818–824 (2009).
This article is an in-depth investigation of how interface morphology affects polymer solar-cell performance.
13. Koleilat, G. I. *et al.* Efficient, stable infrared photovoltaics based on solution-cast colloidal quantum dots. *ACS Nano* **2,** 833–840 (2008).
14. Clifford, J. P. *et al.* Fast, sensitive and spectrally tuneable colloidal-quantum-dot photodetectors. *Nature Nanotechnol.* **4,** 40–44 (2009).
15. Barkhouse, D. A. R. *et al.* Depleted bulk heterojunction colloidal quantum dot photovoltaics. *Adv. Mater.* **23,** 3134–3138 (2011).
16. Kramer, I. J. *et al.* Ordered nanopillar structured electrodes for depleted bulk heterojunction colloidal quantum dot solar cells. *Adv. Mater.* **24,** 2315–2319 (2012).
17. Persson, P., Lundqvist, M. J., Ernstorfer, R., Goddard, W. A. III & Willig, F. Quantum chemical calculations of the influence of anchor-cum-spacer groups on femtosecond electron transfer times in dye-sensitized semiconductor nanocrystals. *J. Chem. Theory Comput.* **2,** 441–451 (2006).
18. Yella, A. *et al.* Porphyrin-sensitized solar cells with cobalt (II/III)-based redox electrolyte exceed 12 percent efficiency. *Science* **334,** 629–634 (2011).
19. Bisquert, J. Theory of the impedance of electron diffusion and recombination in a thin layer. *J. Phys. Chem. B* **106,** 325–333 (2001).
20. Wang, Q. *et al.* Characteristics of high efficiency dye-sensitized solar cells. *J. Phys. Chem. B* **110,** 25210–25221 (2006).
21. Villanueva-Cab, J., Wang, H., Oskam, G. & Peter, L. M. Electron diffusion and back reaction in dye-sensitized solar cells: the effect of nonlinear recombination kinetics. *J. Phys. Chem. Lett.* **1,** 748–751 (2010).
22. Fabregat-Santiago, F. *et al.* Electron transport and recombination in solid-state dye solar cell with spiro-OMeTAD as hole conductor. *J. Am. Chem. Soc.* **131,** 558–562 (2009).
23. Mora-Seró, I. *et al.* Recombination in quantum dot sensitized solar cells. *Acc. Chem. Res.* **42,** 1848–1857 (2009).
24. Klem, E. J. D., MacNeil, D. D., Cyr, P. W., Levina, L. & Sargent, E. H. Efficient solution-processed infrared photovoltaic cells: planarized all-inorganic bulk heterojunction devices via inter-quantum-dot bridging during growth from solution. *Appl. Phys. Lett.* **90,** 183113 (2007).
25. Johnston, K. W. *et al.* Schottky-quantum dot photovoltaics for efficient infrared power conversion. *Appl. Phys. Lett.* **92,** 151115 (2008).
26. Johnston, K. W. *et al.* Efficient Schottky quantum-dot photovoltaics: the roles of depletion, drift, and diffusion. *Appl. Phys. Lett.* **92,** 122111 (2008).
27. Pattantyus-Abraham, A. G. *et al.* Depleted-heterojunction colloidal quantum dot solar cells. *ACS Nano* **4,** 3374–3380 (2010).
This article reports the first colloidal quantum-dot solar cell to exceed 5% solar power conversion efficiency and the architecture that has allowed its subsequent advances.
28. Liu, H. *et al.* Electron acceptor materials engineering in colloidal quantum dot solar cells. *Adv. Mater.* **23,** 3832–3837 (2011).
29. Hyun, B. R. *et al.* Electron injection from colloidal PbS quantum dots into titanium dioxide nanoparticles. *ACS Nano* **2,** 2206–2212 (2008).
30. Koleilat, G. I. *et al.* A donor-supply electrode (DSE) for colloidal quantum dot photovoltaics. *Nano Lett.* **11,** 5173–5178 (2011).
31. Minemoto, T. *et al.* Cu(In,Ga)Se$_2$ solar cells with controlled conduction band offset of window/Cu(In,Ga)Se$_2$ layers. *J. Appl. Phys.* **89,** 8327–8330 (2001).
32. Haight, R. *et al.* Band alignment at the Cu$_2$ ZnSn (S,Se$_{1-x}$)$_4$ /CdS interface. *Appl. Phys. Lett.* **98,** 253502 (2011).
33. Barkhouse, D. A. R. *et al.* Cd-free buffer layer materials on Cu$_2$ZnSn(S,Se$_{1-x}$)$_4$: Band alignments with ZnO, ZnS, and In$_2$S$_3$. *Appl. Phys. Lett.* **100,** 193904–193905 (2012).
34. Wada, T., Kohara, N., Nishiwaki, S. & Negami, T. Characterization of the Cu(In,Ga)Se$_2$/Mo interface in CIGS solar cells. *Thin Solid Films* **387,** 118–122 (2001).
35. Rockett, A. The effect of Na in polycrystalline and epitaxial single-crystal CuIn $_{1-x}$Ga$_x$Se$_2$. *Thin Solid Films* **480–481,** 2–7 (2005).
36. Brabec, C. J. *et al.* Tracing photoinduced electron transfer process in conjugated polymer/fullerene bulk heterojunctions in real time. *Chem. Phys. Lett.* **340,** 232–236 (2001).
37. Veldman, D. *et al.* Compositional and electric field dependence of the dissociation of charge transfer excitons in alternating polyfluorene copolymer/fullerene blends. *J. Am. Chem. Soc.* **130,** 7721–7735 (2008).
38. Vandewal, K., Tvingstedt, K., Gadisa, A., Inganäs, O. & Manca, J. V. On the origin of the open-circuit voltage of polymer-fullerene solar cells. *Nature Mater.* **8,** 904–909 (2009).
39. Gélinas, S. *et al.* The binding energy of charge-transfer excitons localized at polymeric semiconductor heterojunctions. *J. Phys. Chem. C* **115,** 7114–7119 (2011).
40. Lee, J. *et al.* Charge transfer state versus hot exciton dissociation in polymer-fullerene blended solar cells. *J. Am. Chem. Soc.* **132,** 11878–11880 (2010).
41. Maturová, K., Janssen, R. A. J. & Kemerink, M. Connecting scanning tunneling spectroscopy to device performance for polymer:fullerene organic solar cells. *ACS Nano* **4,** 1385–1392 (2010).
42. He, Z. *et al.* Simultaneous enhancement of open-circuit voltage, short-circuit current density, and fill factor in polymer solar cells. *Adv. Mater.* **23,** 4636–4643 (2011).
43. O'Malley, K. M., Li, C.-Z., Yip, H.-L. & Jen, A. K. Y. Enhanced open-circuit voltage in high performance polymer/fullerene bulk-heterojunction solar cells by cathode modification with a C_{60} surfactant. *Adv. Energy Mater.* **2,** 82–86 (2012).
44. Irwin, M. D., Buchholz, D. B., Hains, A. W., Chang, R. P. H. & Marks, T. J. p-Type semiconducting nickel oxide as an efficiency-enhancing anode interfacial layer in polymer bulk-heterojunction solar cells. *Proc. Natl Acad. Sci. USA* **105,** 2783–2787 (2008).
45. Murray, I. P. *et al.* Graphene oxide interlayers for robust, high-efficiency organic photovoltaics. *J. Phys. Chem. Lett.* **2,** 3006–3012 (2011).
46. Gilot, J., Wienk, M. M. & Janssen, R. A. J. Double and triple junction polymer solar cells processed from solution. *Appl. Phys. Lett.* **90,** 143512 (2007).
47. Jin, Y. K. *et al.* Efficient tandem polymer solar cells fabricated by all-solution processing. *Science* **317,** 222–225 (2007).
48. McDonald, S. A. *et al.* Solution-processed PbS quantum dot infrared photodetectors and photovoltaics. *Nature Mater.* **4,** 138–142 (2005).
49. Konstantatos, G. *et al.* Ultrasensitive solution-cast quantum dot photodetectors. *Nature* **442,** 180–183 (2006).
50. Konstantatos, G., Clifford, J., Levina, L. & Sargent, E. H. Sensitive solution-processed visible-wavelength photodetectors. *Nature Photonics* **1,** 531–534 (2007).
51. Barkhouse, D. A. R., Pattantyus-Abraham, A. G., Levina, L. & Sargent, E. H. Thiols passivate recombination centers in colloidal quantum dots leading to enhanced photovoltaic device efficiency. *ACS Nano* **2,** 2356–2362 (2008).
52. Tang, J. *et al.* Colloidal-quantum-dot photovoltaics using atomic-ligand passivation. *Nature Mater.* **10,** 765–771 (2011).
53. Jeong, K. S. *et al.* Enhanced mobility-lifetime products in PbS colloidal quantum dot photovoltaics. *ACS Nano* **6,** 89–99 (2011).
54. Nagpal, P. & Klimov, V. I. Role of mid-gap states in charge transport and photoconductivity in semiconductor nanocrystal films. *Nature Commun.* **2,** 486, http://dx.doi.org/10.1038/ncomms1492 (2011).
55. Hibberd, C. J. *et al.* Non-vacuum methods for formation of Cu(In, Ga)(Se, $S)_2$ thin film photovoltaic absorbers. *Prog. Photovolt. Res. Appl.* **18,** 434–452 (2010).
56. Katagiri, H. *et al.* Development of CZTS-based thin film solar cells. *Thin Solid Films* **517,** 2455–2460 (2009).
57. Scragg, J. J., Ericson, T., Kubart, T., Edoff, M. & Platzer-Björkman, C. Chemical insights into the instability of Cu$_2$ZnSnS$_4$ films during annealing. *Chem. Mater.* **23,** 4625–4633 (2011).
58. Yuan, M. *et al.* Antimony assisted low-temperature processing of CuIn$_{1-x}$Ga$_x$Se$_2$ $_{-y}$S$_y$ solar cells. *Thin Solid Films* **519,** 852–856 (2010).
59. Liao, D. & Rockett, A. Cd doping at the CuInSe$_2$/CdS heterojunction. *J. Appl. Phys.* **93,** 9380–9382 (2003).
60. Rusu, M. *et al.* Three-dimensional structure of the buffer/absorber interface in CdS/ CuGaSe$_2$ based thin film solar cells. *Appl. Phys. Lett.* **95,** 173502 (2009).
61. Hetzer, M. J. *et al.* Direct observation of copper depletion and potential changes at copper indium gallium diselenide grain boundaries. *Appl. Phys. Lett.* **86,** 162105 (2005).
62. Mitzi, D. B. *et al.* Hydrazine-based deposition route for device-quality CIGS films. *Thin Solid Films* **517,** 2158–2162 (2009).
63. Chen, S., Gong, X. G., Walsh, A. & Wei, S. H. Defect physics of the kesterite thin-film solar cell absorber Cu$_2$ ZnSnS$_4$. *Appl. Phys. Lett.* **96,** 021902 (2010).
64. Cao, Q. *et al.* Defects in Cu(In,Ga)Se$_2$ chalcopyrite semiconductors: a comparative study of material properties, defect states, and photovoltaic performance. *Adv. Energy Mater.* **1,** 845–853 (2011).
65. Guha, S. *et al.* Structural, defect, and device behavior of hydrogenated amorphous Si near and above the onset of microcrystallinity. *Appl. Phys. Lett.* **74,** 1860–1862 (1999).
66. Heath, J. T., Cohen, J. D. & Shafarman, W. N. Bulk and metastable defects in CuIn$_{1-x}$Ga$_x$Se$_2$ thin films using drive-level capacitance profiling. *J. Appl. Phys.* **95,** 1000–1010 (2004).
67. Han, L. *et al.* High-efficiency dye-sensitized solar cell with a novel co-adsorbent. *Energy Environ. Sci.* **5,** 6057–6060 (2012).
68. Feldt, S. M. *et al.* Design of organic dyes and cobalt polypyridine redox mediators for high-efficiency dye-sensitized solar cells. *J. Am. Chem. Soc.* **132,** 16714–16724 (2010).
69. Dou, L. *et al.* Tandem polymer solar cells featuring a spectrally matched low-bandgap polymer. *Nature Photonics* **6,** 180–185 (2012).
70. Green, M. A., Emery, K., Hishikawa, Y., Warta, W. & Dunlop, E. D. Solar cell efficiency tables (version 39). *Prog. Photovolt. Res. Appl.* **20,** 12–20 (2012).
71. Luther, J. M. *et al.* Schottky solar cells based on colloidal nanocrystal films. *Nano Lett.* **8,** 3488–3492 (2008).
72. Wang, X. *et al.* Tandem colloidal quantum dot solar cells employing a graded

recombination layer. *Nature Photonics* **5**, 480–484 (2011).
This article reports the first optimized tandem cell to use quantum size tuning alone to achieve the desired bandgaps in the constituent junctions.
73. Todorov, T. K., Reuter, K. B. & Mitzi, D. B. High-efficiency solar cell with earth-abundant liquid-processed absorber. *Adv. Mater.* **22**, E156–E159 (2010).
74. Guo, Q., Ford, G. M., Hillhouse, H. W. & Agrawal, R. in *Proc. 37th IEEE Photovoltaic Specialist Conf.* 003522–003526. (IEEE, 2011).
75. Tang, C. W. Two-layer organic photovoltaic cell. *Appl. Phys. Lett.* **48**, 183–185 (1986).
76. Ma, W., Yang, C., Gong, X., Lee, K. & Heeger, A. J. Thermally stable, efficient polymer solar cells with nanoscale control of the interpenetrating network morphology. *Adv. Funct. Mater.* **15**, 1617–1622 (2005).
77. Park, S. H. *et al.* Bulk heterojunction solar cells with internal quantum efficiency approaching 100%. *Nature Photon.* **3**, 297–302 (2009).
78. Kapoor, V. K., Fisher, M. & Roe, R. *Mat. Res. Soc. Symp. Proc.* H261–H267 (Material Research Society, 2001).
79. Van Duren, J. K. J., Leidholm, C., Pudov, A., Robinson, M. R. & Roussillon, Y. *Materials Research Society Symp. Proc.* 259–268 (Material Research Society, 2002).
80. Todorov, T. K., Gunawan, O., Gokmen, T. & Mitzi, D. B. Solution-processed $Cu(In,Ga)(S,Se)_2$ absorber yielding a 15.2% efficient solar cell. *Prog. Photovolt. Res. Appl.* http://dx.doi.org/10.1002/pip.1253 (2012).
81. Barkhouse, D. A. R., Gunawan, O., Gokmen, T., Todorov, T. K. & Mitzi, D. B. Device characteristics of a 10.1% hydrazine-processed $Cu_2ZnSn(Se,S)_4$ solar cell. *Prog. Photovolt. Res. Appl.* **20**, 6–11 (2012).
82. Arakawa, H. Y. T., Okada, K., Matsui, K., Kitamura, T., Tanabe, N. Highly durable dye-sensitized solar cells. *Fujikura Tech. Rev.* **2009**, 55–59 (2009).
83. Semonin, O. E. *et al.* Peak external photocurrent quantum efficiency exceeding 100% via MEG in a quantum dot solar cell. *Science* **334,** 1530–1533 (2011).
84. Gunawan, O., Todorov, T. K. & Mitzi, D. B. Loss mechanisms in hydrazine-processed Cu_2 ZnSn $(Se,S)_4$ solar cells. *Appl. Phys. Lett.* **97**, 233506 (2010).
85. Veldman, D., Meskers, S. C. J. & Janssen, R. A. J. The energy of charge-transfer states in electron donor-acceptor blends: insight into the energy losses in organic solar cells. *Adv. Funct. Mater.* **19,** 1939–1948 (2009).
86. Voznyy, O. Mobile surface traps in CdSe nanocrystals with carboxylic acid ligands. *J. Phys. Chem. C* **115**, 15927–15932 (2011).
87. Puzder, A., Williamson, A. J., Gygi, F. & Galli, G. Self-healing of CdSe nanocrystals: first-principles calculations. *Phys. Rev. Lett.* **92,** 217401 (2004).
88. Yu, M. *et al.* First principles study of CdSe quantum dots: stability, surface unsaturations, and experimental validation. *Appl. Phys. Lett.* **88,** 231910 (2006).
89. Frenzel, J., Joswig, J. O., Sarkar, P., Seifert, G. & Springborg, M. The effects of organisation, embedding and surfactants on the properties of cadmium chalcogenide (CdS, CdSe and CdS/CdSe) semiconductor nanoparticles. *Eur. J. Inorg. Chem.* **2005,** 3585–3596 (2005).
90. Yablonovitch, E. & Cody, G. D. Intensity enhancement in textured optical sheets for solar cells. *IEEE Trans. Electron. Dev.* **29,** 300–305 (1982).
91. Atwater, H. A. & Polman, A. Plasmonics for improved photovoltaic devices. *Nature Mater.* **9,** 205–213 (2010).
92. Ferry, V. E. *et al.* Optimized spatial correlations for broadband light trapping nanopatterns in high efficiency ultrathin film a-Si:H solar cells. *Nano Lett.* **11,** 4239–4245 (2011).
93. Pillai, S., Catchpole, K. R., Trupke, T. & Green, M. A. Surface plasmon enhanced silicon solar cells. *J. Appl. Phys.* **101,** 093105 (2007).
94. Yang, J. *et al.* Plasmonic polymer tandem solar cell. *ACS Nano* **5,** 6210–6217 (2011).
95. Brown, M. D. *et al.* Plasmonic dye-sensitized solar cells using core-shell metal-insulator nanoparticles. *Nano Lett.* **11,** 438–445 (2011).
96. Klem, E. J. D. *et al.* Planar PbS quantum dot/C_{60} heterojunction photovoltaic devices with 5.2% power conversion efficiency. *Appl. Phys. Lett.* **100,** 173109 (2012).

Acknowledgements. E.H.S acknowledges that this Review is based, in part, on work supported by an award (no. KUS-11-009-21) made by King Abdullah University of Science and Technology (KAUST), by the Ontario Research Fund Research Excellence Program and by the Natural Sciences and Engineering Research Council (NSERC) of Canada. E.H.S acknowledges the contribution of I. Kramer, S. Thon and O. Voznyy to the figures and text. M.G. acknowledges that this Review is based, in part, on work supported by the Stanford University Center of Advanced Molecular Photovoltaics (CAMP) under an award (no. KUS-C1-015-21) made by KAUST and by the European Research Council (ERC) under the Advanced Research Grant No. 247404 (Mesolight project).

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/jvsuoz. Correspondence should be addressed to E.H.S. (ted.sargent@utoronto.ca).

REVIEW

doi:10.1038/nature11477

Membrane-based processes for sustainable power generation using water

Bruce E. Logan[1] & Menachem Elimelech[2]

Water has always been crucial to combustion and hydroelectric processes, but it could become the source of power in membrane-based systems that capture energy from natural and waste waters. Two processes are emerging as sustainable methods for capturing energy from sea water: pressure-retarded osmosis and reverse electrodialysis. These processes can also capture energy from waste heat by generating artificial salinity gradients using synthetic solutions, such as thermolytic salts. A further source of energy comes from organic matter in waste waters, which can be harnessed using microbial fuel-cell technology, allowing both wastewater treatment and power production.

A remarkable amount of energy is available from the salinity difference between sea water and fresh water. In theory, up to 0.8 kilowatts per cubic metre could be extracted — equivalent to the energy generated from water falling over a dam more than 280 metres high[1,2]. The limiting factor in obtaining this energy is the supply of fresh water: about 2 terawatts (1 TW is equal to 1,000 gigawatts) is available globally from rivers flowing into the sea, of which perhaps 980 GW could be harnessed[3]. In addition, wastewater release into the ocean could provide another 18 GW of salinity-gradient power. Although 800 GW of power is currently obtained from hydroelectric processes globally, salinity-gradient energy remains a large and untapped resource. Capturing this energy will require the engineering and development of efficient energy-conversion technologies.

Several approaches to capture salinity-gradient energy are being developed, but the most promising are pressure-retarded osmosis (PRO)[4,5] and reverse electrodialysis (RED)[6,7]. Both are relatively close to commercialization, but their application is limited by cost and fouling of membranes, reducing the useful lifetimes of the membrane-packed modules that both systems use. PRO uses the flow of water — but not ions — through the membranes to produce pressurized water that generates electricity using mechanical turbines. RED uses membranes for ion — but not water — transport, and the electrical current generated is captured directly from the flow of ions. Capacitive systems, which are based on alternate charging (using salt water) and discharging (using fresh water) of materials, may also contain membranes[8–10], but are less advanced than either RED or PRO and are not covered in this Review. PRO and RED have been used mainly to capture natural salinity-gradient energy using sea water and river water[3], but the use of thermolytic solutions offers opportunities to capture the energy from waste heat using these systems[11–13].

Microbial fuel-cell (MFC) technologies can generate energy from organic matter in waste waters and biomass[14]. Microorganisms oxidize the organic matter and release electrons outside the cell, allowing the generation of an electrical current that can be used for producing electrical power, biofuels and valuable chemical products[15]. In the United States, as much as 3% of the electricity generated (15 GW) is used for wastewater treatment[16]. If that energy was no longer needed for treatment, three times that amount could be saved in primary energy (assuming a 33% efficient power plant), with further energy produced from the organic matter in the waste water. The same MFC technologies can also be applied to capture energy from biomass, a resource that could contribute more than 600 GW of renewable power[17]. MFCs can be combined with other processes, such as fermentation of complex organic matter into soluble organics[18,19], and a 'stack' of RED membranes to increase efficiency and to recover additional energy[20,21].

PRO, RED and MFCs are distinct from the more common methods of electricity generation, which are based on harvesting non-renewable materials from nature. These newer processes offer the opportunity to generate energy from abundant but largely unused resources; however, they are at various stages of transfer from the laboratory bench to practical application. We review these three water-based processes and the challenges for their widespread application.

Power production using PRO

PRO extracts salinity-gradient energy by using semipermeable membranes to allow the transport of water from a low-concentration solution (such as river, brackish or waste water) into a high-concentration draw solution (sea water)[3,22–25] (Fig. 1). Sea water can also be used as the low-concentration solution with brines produced from seawater desalination as the draw solution. In theory, the maximum extractable energy during the reversible mixing of a dilute stream with saline draw solutions is substantial, ranging from 0.75 kW-hours per cubic metre to 14.1 kWh per cubic metre of the low-concentration stream (Fig. 1). The actual energy extracted will always be lower because of inherent, irreversible energy losses.

The development of PRO has been hindered for many years by the lack of a membrane capable of allowing an adequate flow[22,23,26,27]. The bulky support layers of the reverse-osmosis membranes cause severe internal concentration polarization (ICP), which markedly reduces permeate-water flux and therefore power density[22,26,28–30]. Cellulose-acetate membranes specifically designed for forward osmosis have substantially reduced ICP effects, stimulating a resurgence in PRO research and development[22,23,31]. The power density has increased from less than 1 W m^{-2} to 2.7 W m^{-2} for river water and sea water, and, using a sea water reverse-osmosis brine[22,31,32], the power density reached 4–5.1 W m^{-2}. Thin-film composite polyamide membranes may allow 5.7–10 W m^{-2} to be achieved because of their higher intrinsic water permeabilities and lower ICP than cellulose-acetate membranes[5,30,33].

The main challenges for implementing PRO for the economical production of energy using natural waters are the development of

[1]Department of Civil and Environmental Engineering, 212 Sackett Building, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. [2]Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520-8286, USA.

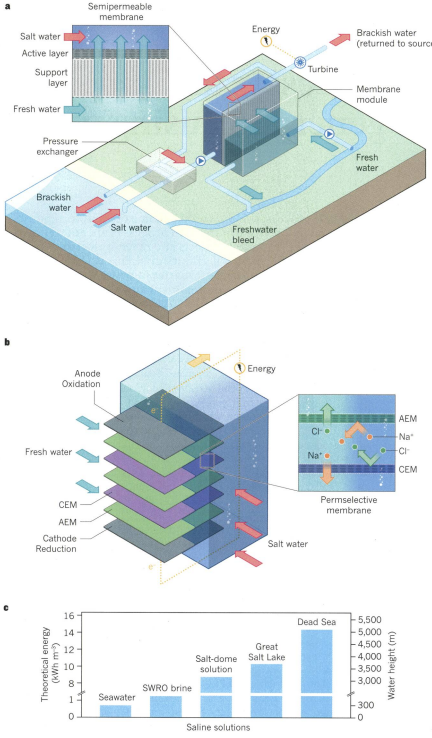


Figure 1 | Comparison of PRO and RED systems. **a**, In PRO, the low-concentration feed solution and high-concentration draw solution are separated by a semipermeable membrane through which, ideally, only water moves from low to high concentration, with ions selectively retained at the membrane active layer. As the two solutions flow in adjacent chambers, water is drawn into the saltwater side across the membrane, diluting the solution and increasing the volume flow rate of water. The resultant high-pressure solution is split into a flow that drives a turbine and a flow that returns to the pressure exchanger. The exchanger transfers pressure energy from a high-pressure fluid stream to a low-pressure fluid stream, thereby pressuring the saltwater feed. The diluted water returns to the source, and the turbine generates energy. **b**, In RED, the energy generated by the mixing of fresh (blue arrows) and salt water (red arrows) is captured using alternating cation-exchange membranes (CEMs) and anion-exchange membranes (AEMs), with low- and high-concentration solution flowing through each alternate channel. These permselective membranes transfer cations through the CEM and anions through the AEM from high- to low-concentration solutions. The difference in electrochemical potential as a result of the positive ions moving one way and the negative ions moving the other is turned into an electrical current at the electrodes, which can be used to generate energy. **c**, The maximum energy that is theoretically extractable from the reversible mixing of a dilute stream with saline solutions from five sources. The height needed to produce the same energy from falling water is also shown. The maximum theoretical energy per cubic metre of the dilute solution is about equal to the osmotic pressure of the respective saline solutions: 27 bar for typical sea water[81], 375 bar for the Great Salt Lake in Utah[82], 507 bar for the Dead Sea on the border of Israel and Jordan[82,83], seawater reverse-osmosis plant (SWRO) brine was assumed to be 54 bar (for 50% recovery) and 316 bar for salt-dome (subsurface geological structures) solution assumed to be 5 M sodium chloride[82] (calculated using OLI Stream Analyzer software).

low-cost and robust membranes that have minimal ICP and fouling and the potential environmental impacts caused by the disruption of the natural flow of water. In addition, feed and draw streams need to be extensively pretreated to prevent membrane fouling and deterioration in performance potential. A plant in Norway, used to pilot PRO as a method of generating power from natural salinity gradients, generated less than 1 W m^{-2} using commercial, asymmetrical cellulose-acetate membranes[23]. This is much lower than the target power density of 5 W m^{-2} needed to make PRO economically viable[23,27,34]. Pretreatment of river water and sea water can consume a considerable amount of energy (0.2–0.3 kWh m^{-3}) (ref. 35). In a PRO plant with a 50% overall efficiency, the actual extractable energy from the mixing of river water and sea water is 0.3–0.4 kWh m^{-3}, leaving about 0.1 kWh of useful energy that can be derived per cubic metre of river-water feed solution. This energy consumption indicates the need to develop fouling-resistant membranes with tailored surface properties and membrane modules with improved hydrodynamic mixing to mitigate fouling.

Electricity production using RED

The advantage of using a RED system for capturing energy is that electricity is generated directly from salinity gradients. Sea water and fresh water are introduced into arrays or stacks of membranes with alternating anion-exchange membranes and cation-exchange membranes, directly generating an electrochemical potential (Fig. 1). In theory, water can be split using about 10 membrane pairs because each pair generates around 0.1–0.2 V from typical river and sea water. But in

reality, the energy needed to overcome energy losses at the electrodes, or overpotential, means that at least 20 pairs are needed.

Alternatives to water splitting have been investigated to increase energy capture and improve safety. Water splitting releases oxygen and, from seawater, toxic chlorine gas from the anode, and potentially explosive hydrogen gas at the cathode that needs to be either used properly or vented. Iron-based oxidation–reduction (redox) couples could allow recycling of electrolyte solutions, avoiding the need to split water[36]. Ferrocyanide–ferricyanide has been used as a reversible redox couple, but it could potentially produce toxic hydrogen-cyanide gas at the electrodes. In addition, some leakage of these compounds into natural waters is inevitable because of the non-ideal permselectivity of ion-exchange membranes. The ferrous–ferric iron redox couple is also promising, but iron can precipitate on and within membranes and electrode surfaces, leading to fouling and a decrease in current. Although there are energy losses at the electrodes, when 50–100 membrane pairs are used to generate electrical power, these losses become an increasingly smaller percentage of the overall energy balance, and are therefore acceptable.

Substantial advances are being made to increase RED power densities and energy efficiencies through improvement of membrane materials, spacing and architecture. The highest power density for a RED stack with 50 cell pairs is 0.93 W m^{-2} total membrane area at about 3 V, neglecting electrode overpotentials and pumping losses (based on each type of ion-exchange membrane having a 0.5 m^2 total surface area, or 93 W m^{-2} based on the planar cross-sectional area between the electrodes)[37]. Pumping required roughly 25% of the total power produced[37]. A larger stack (with a total area of 19 m^2) generated a net power of 0.4 W m^{-2}, with an energy efficiency of 18%, although efficiency could have been increased to 30% with a lower net power density (0.35 W m^{-2}) (ref. 7). In 25-membrane-pair stacks, optimizing membrane materials and reducing the spacing between membranes increased power densities to 1.2–2.2 W m^{-2} (refs 38, 39). The use of ion-conductive spacers between the membranes[40] or introducing ridges and flow patterns into the membrane material to avoid using spacers[41] are promising methods for reducing the size of RED systems and increasing power generation. Power density is estimated to reach to 3.4–4 W m^{-2} for RED[3,39] and 7.7 W m^{-2} (4 kW m^{-3}) for PRO when river–seawater solutions are used[3]. For these estimated power densities, 9 W reactors with comparable power per area would produce 0.16 kW m^{-3} for RED, and 1.1 kW m^{-3} for PRO[3,38,42].

The main challenge for the commercialization of RED is the cost of ion-exchange membranes, but a global increase in demand could lower costs. Over the past decade, new materials, improved fabrication methods and increased production of reverse-osmosis membranes resulted in a decrease in the cost, which was a major factor in the increased use of reverse osmosis for water desalination[16].

Closed-loop PRO or RED using synthetic solutions

Open-loop PRO and RED use natural water or wastewater effluents that require extensive pretreatment and fouling-control measures, and they can have an environmental impact because of the withdrawal of natural water from ecosystems. Closed-loop systems use synthetic solutions, which avoid these problems and potentially produce much higher power densities. A closed-loop system that converts thermal energy to mechanical work or electricity — termed a heat engine[22,43] — is ideally suited to sustainable energy production (Fig. 2). In such a system, low- and high-concentration streams are used to generate energy by PRO or RED, and thermal energy is used to separate the resultant mixed stream into low- and high-concentration streams that are recycled back into the system. The heat engine can be economically viable if waste heat is used in the separation stage; this low-grade heat can come from industrial processes or power plants, or as geothermal energy from shallow wells[44].

Closed-loop systems use salt solutions that can be separated by low-grade heat. Of these thermolytic solutions, an ammonia–carbon-dioxide solution is particularly promising and is capable of generating high power densities of 250 W m^{-2} at 5 M concentrations (Fig. 2)[12,44]. Very soluble ammonium salts can create high osmotic pressures, are rejected

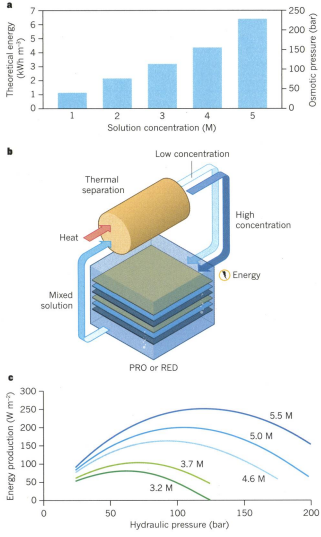


Figure 2 | A heat engine. a, The maximum energy theoretically extractable based on the reversible mixing of ammonia–carbon-dioxide solutions with water. The osmotic pressures of 1–5 M solutions show that there is high potential for energy generation compared with open-loop PRO systems that use sea water and river water. **b**, A closed-loop system in which synthetic low- and high-concentration solutions pass through a PRO or RED system to generate electrical power. Thermal energy is used to separate the mixed solution back into high- and low-concentration streams, which are then recycled back into the system. **c**, Theoretical membrane power densities achievable by the ammonia–carbon-dioxide osmotic heat engine for 3.2–5.5 M solutions at various applied hydraulic pressures. Maximum power density is achieved with a hydraulic pressure that is roughly 50% of the ammonia–carbon-dioxide solution osmotic pressure. Data used in parts **a** and **c** taken from McGinnis *et al.*[12].

well by semipermeable membranes and can be separated from water and recycled with low-grade heat by conventional distillation processes[12,45]. In theory, a large amount of energy can be extracted from the reversible mixing of a dilute stream with the ammonia–carbon-dioxide solution (Fig. 2). Using low-grade heat at 50 °C, the thermal energy needed to separate the diluted ammonia–carbon-dioxide solution into water (low-concentration stream), and ammonia and carbon dioxide (that reform the high-concentration stream) is 358 MJ m^{-3} (99 kWh m^{-3}) for a 1 M solution and 593 MJ m^{-3} (165 kWh m^{-3}) for a 2 M solution[12].

Highly soluble inorganic salts, such as sodium chloride, can also be used in closed-loop heat engines with separation processes that use low-grade heat. For example, membrane distillation uses a vapour pressure gradient created by the temperature difference across the membrane to transport water vapour through microporous hydrophobic membranes[46]. Thermal-energy requirements for membrane distillation

can range from 40 kWh m^{-3} to 650 kWh m^{-3}, depending on membrane properties, system design, heat source and operating temperature[47].

To successfully implement closed-loop systems that use synthetic solutions, several challenges need to be addressed. The main challenge for PRO is the development of membranes with sufficient mechanical strength to withstand the high hydraulic pressures that occur with very high salt concentrations (Fig. 2). For membrane distillation to be viable in these systems, efficient and low-cost membranes need to be developed.

Energy production from waste matter using MFCs

Waste water can be a substantial source of energy. In the United Kingdom, the energy content of waste waters ranges from 2.1 kWh m^{-3} for domestic to 4.7 kWh m^{-3} for mixed-use (commercial and domestic) water[48]. Typical wastewater treatment plants use high concentrations of bacteria supplied with oxygen, through aeration of the water, to remove the organic matter. The overall energy consumption in conventional plants with this treatment is roughly 0.6 kWh m^{-3} with about half of that used for aeration[16]. Some treatment plants have become energy neutral, mostly though the production of methane gas from solids captured or produced during treatment[16].

MFCs offer a unique approach to directly capture the energy in water, effectively turning a wastewater treatment plant into a power plant[49,50]. Exoelectrogenic bacteria release protons into the waste water and transfer electrons to electrically conductive, inert and high porosity anodes — such as carbon cloth, felt and graphite fibre brushes[50,51] (Fig. 3). At the cathode, the electrons and protons combine with oxygen to form water, generating an overall maximum voltage of about 0.5 V, compared with a theoretical maximum of 1.1 V (ref. 52). In this process, oxygen is transferred passively from the air to the cathode, not through water aeration, providing substantial energy savings compared with conventional processes, even without including the energy that would have been needed for the production of power. For waste waters, power generation using MFCs is usually limited by low solution conductivity (1 mS cm^{-1}), organic-matter concentration, temperature, and microbial kinetics and rates of oxygen reduction under these conditions. Under optimal conditions with acetate as a fuel, and using very large cathodes relative to the projected area of the anode, power densities have reached 6.9 W m^{-2} (normalized to the cathode surface area)[53] and 1.55 kW per cubic metre of reactor volume (2.77 W m^{-2}) (ref. 54). That is considerably higher than that typically produced using actual waste waters (less than 0.3 W m^{-2}), although 1.24 W m^{-2} (0.41 W m^{-3} based on liquid volume) has been achieved using special carbon-nanotube-coated sponge anodes[55]. There have been few tests on reactors larger than tens of litres[56–58] and none of systems larger than one cubic metre. Power densities have remained relatively constant with increased reactor size when the area of the cathode per unit volume of the reactor is kept constant[57,59,60]. Low power densities and a lack of demonstrations of larger-scale systems have led to debate on whether MFC-based treatment plants can ever compete with more conventional wastewater treatment systems that are engineered to achieve a slight net power production[16].

By modifying MFCs, they can also be used for chemical remediation and to produce hydrogen, methane and acetate. For example, thermodynamically favourable reactions can be generated at the cathode without oxygen by using electron acceptors such as ferrous iron[61]. Pollutants in waste waters, such as nitrates, copper, chromium and vanadium, can be precipitated or removed in MFC processes, saving energy that would otherwise be used in treatment[62–64]. Adding energy to the potential generated by the anode bacteria can increase the current densities, allowing otherwise non-spontaneous reactions. Adding voltage to that already produced by bacteria (typically –0.25 V or less) reduces that needed to produce hydrogen gas (–0.414 V), although more than this difference is added because of overpotentials[65,66]. Inorganic catalysts, enzymes or microorganisms can be used to reduce these overpotentials. Methane is produced by methanogenic microorganisms from acetate or hydrogen, and it can also be produced from electrical current using methanogens; however, this bioelectrochemical production requires additional voltage to be added to the cathode to overcome electrode overpotentials. Appreciable methane can be produced at cathode potentials more negative than –0.7 V using methanogens on a cathode[67], compared with the –0.24 V that is theoretically needed. Organic chemicals, such as acetate, can also be produced at the cathode directly from electrical currents, using microorganisms[67–69].

Cellulosic and algal biomass sources, among others, can be used in MFCs[18,70,71], but they must be made soluble first, and usually converted by fermentation into simple molecules that the exoelectrogenic bacteria can use — such as acetate, lactate, pyruvate, formate and ethanol. The generation of biofuels from cellulose through enzyme and fermentation processes is well under way, with ethanol and hydrogen as the main products. Bioelectrochemical systems can also be used to produce electricity and biofuels either from using these same sources of organic matter or using waste streams from biorefineries[72,73].

The main challenges for the implementation of MFCs for energy production are the cost and longevity of the electrodes, and achieving high current densities through the reduction of internal resistances by improved stack design. Anodes are relatively inexpensive and are stable over the long term, but the cathodes are not[59,60]. Although the use of precious metals in the cathode can be avoided by using activated carbon for oxygen reduction, the performance drops noticeably over time[74,75].

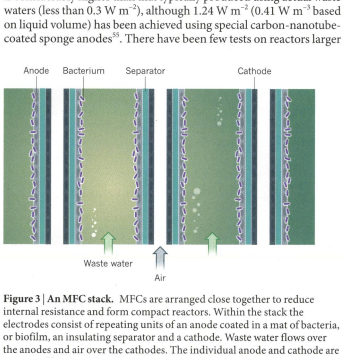


Figure 3 | An MFC stack. MFCs are arranged close together to reduce internal resistance and form compact reactors. Within the stack the electrodes consist of repeating units of an anode coated in a mat of bacteria, or biofilm, an insulating separator and a cathode. Waste water flows over the anodes and air over the cathodes. The individual anode and cathode are connected by a wire (not shown).

Hybrid technologies

Various combinations of technologies are being explored to create more synergistic approaches (Fig. 4). For example, a brine solution could be used as the high-concentration stream in RED or PRO, and sea water as the low-concentration stream[76] (Fig. 4a). Waste water can be used to dilute sea water, with a forward-osmosis membrane either placed inside a conventional bioreactor or used to extract water after wastewater treatment[77]. Because the energy demand of reverse-osmosis desalination decreases with decreasing feed-water salinity, using waste water to dilute seawater reduces the overall energy of sea water desalination.

Forward-osmosis membranes can also be used with various bioreactors, including MFCs. The salinity of waste water increases during treatment with a forward-osmosis membrane[78] as a result of the increased solution conductivity, reducing the internal resistance of the anode — and improving current densities. Placing the cathode in the draw solution eliminates solution resistance in the cathode chamber[79](Fig. 4b). The maximum power density from an MFC using a forward-osmosis membrane (called an osmotic MFC) was 15% higher with a 35 g l^{-1} sodium-chloride solution than from the MFC alone[79]. A substantial concentration

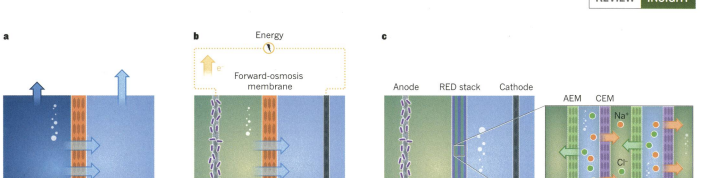


Figure 4 | Hybrid systems could allow further energy recovery. a, Power can be extracted from sea water in PRO or RED systems by using brine as the high-concentration (HC) stream. The two concentrations are separated by either a RED stack or a PRO membrane depending on the system. Either ions (in RED) or water (in PRO) moves from high to low concentration (LC) (see Fig. 1). In PRO, the volume leaving the system is lower for brine than it is for sea water as a result of water flow through the membrane. Once used, the water would be discharged into the sea. **b**, In a forward-osmosis MFC, waste water is used for desalination. Sea water acts as a draw solution, producing a less salty solution through a forward-osmosis membrane and generating a current. **c**, A RED stack built into an MFC increases the voltage and power production compared with an MFC alone. AEM, anion-exchange membrane; CEM, cation-exchange membrane.

of organics in the waste water because of water loss could allow the use of methane- and energy-producing anaerobic digestion, rather than energy-consuming aerobic systems, such as activated sludge. However, if the treated waste water has very high salinities, this could prohibit its discharge into certain environments or its further reuse.

Microbial RED cells are hybrid reactors that have a RED stack placed directly between the MFC electrodes[20,21] (Fig. 4c). When a five-membrane-pair RED stack was used in such a cell for electricity generation, the maximum voltage (1.3 V) and power density (4.3 W m^{-2}) were substantially increased compared with those obtained with the MFC alone (0.5 V, 0.7 W m^{-2}), with acetate as the fuel and sodium-chloride solutions pumped through the RED stack[21]. The use of ammonium bicarbonate in the RED stack further increased performance to 5.6 W m^{-2} with acetate fuel[13]. When domestic waste water was used with ammonium bicarbonate in the RED stack, the maximum power was about 2.8 W m^{-2}, which was nearly an order of magnitude higher than when waste water alone was used[13]. In addition, treatment of the soluble organics was completed within a few hours, which is much shorter than the time needed for domestic wastewater treatment using only the MFC.

Hydrogen gas can be produced in a modified MFC without any external power source by placing the RED stack between the electrodes[20,80]. The energy efficiency achieved from this, based on total energy in the acetate solution and ammonium-bicarbonate solutions entering and leaving the reactor, was 35% (ref. 80). This is lower than the 65% efficiency produced using sodium-chloride solutions with similar concentrations to sea water and river water[20], suggesting that the performance using thermolytic solutions such as ammonium bicarbonate could be increased through system optimization.

Hybrid systems are in an early stage of development, but so far combining technologies seems to be a promising way to overcome the limitations of individual systems. The main challenges, like those for individual systems, are the high cost of materials and achieving high power densities. Hybrid systems will also require co-localization of different energy sources, such as waste water and sea water, or waste water and heat.

Outlook and future directions

Although the advantages of capturing energy using PRO and RED from natural waters are known, relatively little attention has been given to waste heat, brines or waste waters. Conventional power plants provide abundant waste heat — a typical efficiency of 33% would generate 2 GW of waste heat for every 1 GW of electrical power. If we optimistically assume that PRO can recover 60% of the energy in a thermolytic solution such as ammonium bicarbonate, and that 25% of the waste heat from a power plant can be captured in this solution, then PRO could produce as much as 300 MW of additional power (to give an overall energy efficiency of 43%). RED processes may be able to capture only about half this energy, but they can be designed to make use of valuable products at the electrodes, such as using hydrogen as an energy carrier or oxygen for improving combustion efficiency. Energy recovery from brines could be used to improve energy efficiencies of desalination plants, especially in hot, arid environments, where additional low-grade heat could be captured and used in the systems. Although waste water and biomass offer opportunities to improve energy capture, for waste-water treatment, the main benefit may be the energy saved or even just an energy-neutral wastewater treatment process that could allow the adequate sanitation of water globally[16].

The energy-producing processes described in this Review depend on effective and inexpensive membranes that are uniquely tailored to specific solutions and their application. For example, PRO membranes must be optimized for water transport and ion selectivity, and either be resistant to fouling when using sea water and river water or able to remain stable and intact when using thermolytic solutions with high hydraulic pressures. In all applications, power densities need to be high and the cost of membranes very low for these processes to be competitive with other methods of energy production[3]. Tremendous improvements have been made to reverse-osmosis membranes during the past few decades, and membranes for processes driven by osmosis have begun to show substantial improvements, increasing the possible applications of PRO. The cost of ion-exchange membranes could be expected to fall with system advances, which would allow the application of RED[40,41]. For natural waters, a reduction in fouling and improvements to the fouling resistance of membranes are crucial to advancing applications. With these improvements, the energy available from salinity gradients, waste water and biomass could be effectively harnessed as renewable and environmentally sustainable sources of power. ■

1. Yip, N. Y. & Elimelech, M. Thermodynamic and energy efficiency analysis of power generation from natural salinity gradients by pressure retarded osmosis. *Environ. Sci. Technol.* **46**, 5230–5239 (2012).
This paper provides a thorough analysis of the fraction of energy accessible from mixing of fresh river water with sea water for conversion to usable work.
2. Pattle, R. E. Production of electric power by mixing fresh and salt water in the hydroelectric pile. *Nature* **174**, 660 (1954).
3. Ramon, G. Z., Feinberg, B. J. & Hoek, E. M. V. Membrane-based production of

salinity-gradient power. *Energy Environ. Sci.* **4,** 4423–4434 (2011).
4. Lee, K. P., Arnot, T. C. & Mattia, D. A review of reverse osmosis membrane materials for desalination — development to date and future potential. *J. Memb. Sci.* **370,** 1–22 (2011).
5. Chou, S. *et al.* Thin-film composite hollow fiber membranes for pressure retarded osmosis (PRO) process with high power density. *J. Memb. Sci.* **389,** 25–33 (2012).
6. Post, J. W., Hamelers, H. V. M. & Buisman, C. J. N. Energy recovery from controlled mixing salt and fresh water with a reverse electrodialysis system. *Environ. Sci. Technol.* **42,** 5785–5790 (2008).
7. Veerman, J., Saakes, M., Metz, S. J. & Harmsen, G. J. Electrical power from sea and river water by reverse electrodialysis: a first step from the laboratory to a real power plant. *Environ. Sci. Technol.* **44,** 9207–9212 (2010).
8. Brogioli, D. Extracting renewable energy from a salinity difference using a capacitor. *Phys. Rev. Lett.* **103,** 058501 (2009).
9. La Mantia, F., Pasta, M., Deshazer, H. D., Logan, B. E. & Cui, Y. Batteries for efficient energy extraction from a water salinity difference. *Nano Lett.* **11,** 1810–1813 (2011).
10. Sales, B. B. *et al.* Direct power production from a water salinity difference in a membrane-modified supercapacitor flow cell. *Environ. Sci. Technol.* **44,** 5661–5665 (2010).
11. McCutcheon, J. R., McGinnis, R. L. & Elimelech, M. A novel ammonia–carbon dioxide forward (direct) osmosis desalination process. *Desalination* **174,** 1–11 (2005).
12. McGinnis, R. L., McCutcheon, J. R. & Elimelech, M. A novel ammonia–carbon dioxide osmotic heat engine for power generation. *J. Memb. Sci.* **305,** 13–19 (2007).
This article proposes a closed-loop PRO system for power generation using a thermolytic draw solution and low-grade heat.
13. Cusick, R. D., Kim, Y. & Logan, B. E. Energy capture from thermolytic solutions in microbial reverse-electrodialysis cells. *Science* **335,** 1474–1477 (2012).
This article describes a hybrid system that uses MFC technology to provide favourable electrode reactions, allowing the efficient capture of salinity-gradient energy.
14. Logan, B. E. & Regan, J. M. Microbial fuel cells — challenges and applications. *Environ. Sci. Technol.* **40,** 5172–5180 (2006).
15. Logan, B. E. Exoelectrogenic bacteria that power microbial fuel cells. *Nature Rev. Microbiol.* **7,** 375–381 (2009).
16. McCarty, P. L., Bae, J. & Kim, J. Domestic wastewater treatment as a net energy producer — can this be achieved? *Environ. Sci. Technol.* **45,** 7100–7106 (2011).
17. Perlack, R. D. *et al. Biomass as Feedstock for a Bioenergy and Bioproducts Industry: the Technical Feasibility of a Billion-ton Annual Supply* (Oak Ridge National Laboratory, 2005).
18. Lalaurette, E., Thammannagowda, S., Mohagheghi, A., Maness, P.-C. & Logan, B. E. Hydrogen production from cellulose in a two-stage process combining fermentation and electrohydrogenesis. *Int. J. Hydrogen Energy* **34,** 6201–6210 (2009).
19. Lu, L., Ren, N., Xing, D. & Logan, B. E. Hydrogen production with effluent from an ethanol-H_2-coproducing fermentation reactor using a single-chamber microbial electrolysis cell. *Biosens. Bioelectron.* **24,** 3055–3060 (2009).
20. Kim, Y. & Logan, B. E. Hydrogen production from inexhaustible supplies of fresh and salt water using microbial reverse-electrodialysis electrolysis cells. *Proc. Natl Acad. Sci. USA* **108,** 16176–16181 (2011).
21. Kim, Y. & Logan, B. E. Microbial reverse electrodialysis cells for synergistically enhanced power production. *Environ. Sci. Technol.* **45,** 5834–5839 (2011).
22. Achilli, A. & Childress, A. E. Pressure retarded osmosis: From the vision of Sidney Loeb to the first prototype installation – Review. *Desalination* **261,** 205–211 (2010).
23. Nijmeijer, K. & Metz, S. in *Sustainability Science and Engineering* Vol. 2 95–139 (Elsevier, 2010).
24. Thorsen, T. & Holt, T. The potential for power production from salinity gradients by pressure retarded osmosis. *J. Memb. Sci.* **335,** 103–110 (2009).
25. Achilli, A., Cath, T. Y. & Childress, A. E. Selection of inorganic-based draw solutions for forward osmosis applications. *J. Memb. Sci.* **364,** 233–241 (2010).
26. Lee, K. L., Baker, R. W. & Lonsdale, H. K. Membranes for power-generation by pressure-retarded osmosis. *J. Memb. Sci.* **8,** 141–171 (1981).
This paper describes the basis for analysing the performance limiting effects of membranes in PRO.
27. Gerstandt, K., Peinemann, K. V., Skilhagen, S. E., Thorsen, T. & Holt, T. Membrane processes in energy supply for an osmotic power plant. *Desalination* **224,** 64–70 (2008).
28. Cath, T. Y., Childress, A. E. & Elimelech, M. Forward osmosis: Principles, applications, and recent developments. *J. Memb. Sci.* **281,** 70–87 (2006).
29. Yip, N. Y., Tiraferri, A., Phillip, W. A., Schiffman, J. D. & Elimelech, M. High performance thin-film composite forward osmosis membrane. *Environ. Sci. Technol.* **44,** 3812–3818 (2010).
30. Yip, N. Y. *et al.* Thin-film composite pressure retarded osmosis membranes for sustainable power generation from salinity gradients. *Environ. Sci. Technol.* **45,** 4360–4369 (2011).
31. Achilli, A., Cath, T. Y. & Childress, A. E. Power generation with pressure retarded osmosis: an experimental and theoretical investigation. *J. Memb. Sci.* **343,** 42–52 (2009).
32. Zhao, S., Zou, L., Tang, C. Y. & Mulcahy, D. Recent developments in forward osmosis: opportunities and challenges. *J. Memb. Sci.* **396,** 1–21 (2012).
33. Yip, N. Y. & Elimelech, M. Performance limiting effects in power generation from salinity gradients by pressure retarded osmosis. *Environ. Sci. Technol.* **45,** 10273–10282 (2011).
34. Skilhagen, S. E. Osmotic power — a new, renewable energy source. *Desalin. Water Treat.* **15,** 271–278 (2010).
35. Elimelech, M. & Phillip, W. A. The future of seawater desalination: energy, technology, and the environment. *Science* **333,** 712–717 (2011).
36. Veerman, J., Saakes, M., Metz, S. J. & Harmsen, G. J. Reverse electrodialysis: evaluation of suitable electrode systems. *J. Appl. Electrochem.* **40,** 1461–1474 (2010).
This paper reports the production of a very high power density using RED and a large stack of membranes, owing mainly to an optimized cell design.
37. Veerman, J., Saakes, M., Metz, S. J. & Harmsen, G. J. Reverse electrodialysis: performance of a stack with 50 cells on the mixing of sea and river water. *J. Memb. Sci.* **327,** 136–144 (2009).
38. Veerman, J., de Jong, R. M., Saakes, M., Metz, S. J. & Harmsen, G. J. Reverse electrodialysis: comparison of six commercial membrane pairs on the thermodynamic efficiency and power density. *J. Memb. Sci.* **343,** 7–15 (2009).
39. Vermaas, D. A., Saakes, M. & Nijmeijer, K. Doubled power density from salinity gradients at reduced intermembrane distances. *Environ. Sci. Technol.* **45,** 7089–7095 (2011).
This article describes how ion-exchange membranes are configured to also act as spacers, decreasing hydraulic friction and allowing greater flexibility in stack design.
40. Długolecki, P., Dabrowska, J., Nijmeijer, K. & Wessling, M. Ion conductive spacers for increased power generation in reverse electrodialysis. *J. Memb. Sci.* **347,** 101–107 (2010).
41. Vermaas, D. A., Saakes, M. & Nijmeijer, K. Power generation using profiled membranes in reverse electrodialysis. *J. Memb. Sci.* **385–386,** 234–242 (2011).
42. Xu, Y., Peng, X., Tang, C. Y., Fu, Q. S. & Nie, S. Effect of draw solution concentration and operating conditions on forward osmosis and pressure retarded osmosis performance in a spiral wound module. *J. Memb. Sci.* **348,** 298–309 (2010).
43. Loeb, S. Method and apparatus for generating power utilizing pressure-retarded osmosis. US Patent 4,193,267 (1980).
44. McGinnis, R. L. & Elimelech, M. Global challenges in energy and water supply: the promise of engineered osmosis. *Environ. Sci. Technol.* **42,** 8625–8629 (2008).
45. McGinnis, R. L. & Elimelech, M. Energy requirements of ammonia-carbon dioxide forward osmosis desalination. *Desalination.* **207,** 370–382 (2007).
46. Alkhudhiri, A., Darwish, N. & Hilal, N. Membrane distillation: a comprehensive review. *Desal.* **287,** 2–18 (2012).
47. Souhaimi, M. K. & Matsuura, T. *Membrane Distillation: Principles and Applications* (Elsevier, 2011).
48. Heidrich, E. S., Curtis, T. P. & Dolfing, J. Determination of the internal chemical energy of wastewater. *Environ. Sci. Technol.* **45,** 827–832 (2011).
49. Logan, B. E. *et al.* Microbial fuel cells: methodology and technology. *Environ. Sci. Technol.* **40,** 5181–5192 (2006).
50. Logan, B. E. *Microbial Fuel Cells.* (Wiley Blackwell, 2008).
51. Logan, B. E., Cheng, S., Watson, V. & Estadt, G. Graphite fiber brush anodes for increased power production in air-cathode microbial fuel cells. *Environ. Sci. Technol.* **41,** 3341–3346 (2007).
52. Liu, H. & Logan, B. E. Electricity generation using an air-cathode single chamber microbial fuel cell in the presence and absence of a proton exchange membrane. *Environ. Sci. Technol.* **38,** 4040–4046 (2004).
53. Fan, Y., Sharbrough, E. & Liu, H. Quantification of the internal resistance distribution of microbial fuel cells. *Environ. Sci. Technol.* **42,** 8101–8107 (2008).
This paper provides a thorough analysis of the factors that limit power production in MFCs.
54. Fan, Y., Hu, H. & Liu, H. Sustainable power generation in microbial fuel cells using bicarbonate buffer and proton transfer mechanisms. *Environ. Sci. Technol.* **41,** 8154–8158 (2007).
55. Xie, X. *et al.* Carbon nanotube-coated macroporous sponge for microbial fuel cell electrodes. *Energy Environ. Sci.* **5,** 5265–5270 (2012).
56. Dekker, A., Heijne, A. T., Saakes, M., Hamelers, H. V. M. & Buisman, C. J. N. Analysis and improvement of a scaled-up and stacked microbial fuel cell. *Environ. Sci. Technol.* **43,** 9038–9042 (2009).
57. Logan, B. E. Scaling up microbial fuel cells and other bioelectrochemical systems. *Appl. Microbiol. Biotechnol.* **85,** 1665–1671 (2010).
58. Jiang, D. *et al.* A pilot-scale study on utilizing multi-anode/cathode microbial fuel cells (MAC MFCs) to enhance the power production in wastewater treatment. *Int. J. Hydrogen Energy* **36,** 876–884 (2011).
59. Liu, H., Cheng, S., Huang, L. & Logan, B. E. Scale up of a single-chamber microbial fuel cell through optimization of the anode to cathode area ratio. *J. Power Sources* **179,** 274–279 (2008).
60. Cheng, S. & Logan, B. E. Increasing power generation for scaling up single-chamber air cathode microbial fuel cells. *Bioresour. Technol.* **102,** 4468–4473 (2011).
61. ter Heijne, A., Hamelers, H. V. M., de Wilde, V., Rozendal, R. R. & Buisman, C. J. N. Ferric iron reduction as an alternative for platinum-based cathodes in microbial fuel cells. *Environ. Sci. Technol.* **40,** 5200–5205 (2006).
62. Clauwaert, P. *et al.* Biological denitrification in microbial fuel cells. *Environ. Sci. Technol.* **41,** 3354–3360 (2007).
63. Heijne, A. T. *et al.* Copper recovery combined with electricity production in a microbial fuel cell. *Environ. Sci. Technol.* **44,** 4376–4381 (2010).
64. Zhang, B., Feng, C., Ni, J., Zhang, J. & Huang, W. Simultaneous reduction of vanadium (V) and chromium (VI) with enhanced energy recovery based on microbial fuel cell technology. *J. Power Sources* **204,** 34–39 (2012).

65. Liu, H., Grot, S. & Logan, B. E. Electrochemically assisted microbial production of hydrogen from acetate. *Environ. Sci. Technol.* **39,** 4317–4320 (2005).
66. Rozendal, R. A., Hamelers, H. V. M., Euverink, G. J. W., Metz, S. J. & Buisman, C. J. N. Principle and perspectives of hydrogen production through biocatalyzed electrolysis. *Int. J. Hydrogen Energy* **31,** 1632–1640 (2006).
67. Nevin, K. P., Woodard, T. L., Franks, A. E., Summers, A. M. & Lovley, D. R. Microbial electrosynthesis: feeding microbes electricity to convert carbon dioxide and water to multicarbon extracellular organic compounds. *mBio* **1,** e00103–10 (2010).
68. Rabaey, K. & Rozendal, R. A. Microbial electrosynthesis – revisiting the electrical route for microbial production. *Nature Rev. Microbiol.* **8,** 706–716 (2010).
69. Nevin, K. P. *et al.* Electrosynthesis of organic compounds from carbon dioxide is catalyzed by a diversity of acetogenic microorganisms. *Appl. Environ. Microbiol.* **77,** 2882–2886 (2011).
70. Ren, Z., Ward, T. E. & Regan, J. M. Electricity production from cellulose in a microbial fuel cell using a defined binary culture and an undefined mixed culture. *Environ. Sci. Technol.* **41,** 4781–4786 (2007).
71. Velásquez-Orta, S. B., Curtis, T. P. & Logan, B. E. Energy from algae using microbial fuel cells. *Biotechnol. Bioeng.* **103,** 1068–1076 (2009).
72. Pant, D., Bogaert, G. V., Diels, L. & Vanbroekhoven, K. A review of the substrates used in microbial fuel cells (MFCs) for sustainable energy production. *Bioresour. Technol.* **101,** 1533–1543 (2009).
73. Borole, A. P. Improving energy efficiency and enabling water recycling in biorefineries using bioelectrochemical systems. *Biofuels, Bioprod. Bioref.* **5,** 28–36 (2011).
74. Zhang, F., Cheng, S., Pant, D., Bogaert, G. V. & Logan, B. E. Power generation using an activated carbon and metal mesh cathode in a microbial fuel cell. *Electrochem. Commun.* **11,** 2177–2179 (2009).
75. Zhang, F., Pant, D. & Logan, B. E. Long-term performance of activated carbon air cathodes with different diffusion layer porosities in microbial fuel cells. *Biosens. Bioelectron.* **30,** 49–55 (2011).
76. Cath, T. Y., Childress, A. E. & Martinetti, C. R. Combined membrane-distillation-forward-osmosis systems and methods of use. US Patent 8,029,671 (2011).
77. Hoover, L. A., Phillip, W. A., Tiraferri, A., Yip, N. Y. & Elimelech, M. Forward with osmosis: Emerging applications for greater sustainability. *Environ. Sci. Technol.* **45,** 9824–9830 (2011).
78. Lay, W. C. L. *et al.* Study of integration of forward osmosis and biological process: membrane performance under elevated salt environment. *Desalination.* **283,** 123–130 (2011).
79. Zhang, F., Brastad, K. S. & He, Z. Integrating forward osmosis into microbial fuel cells for wastewater treatment, water extraction and bioelectricity generation. *Environ. Sci. Technol.* **45,** 6690–6696 (2011).
80. Nam, J.-Y., Cusick, R. D., Kim, Y. & Logan, B. E. Hydrogen generation in microbial reverse-electrodialysis electrolysis cells using a heat-regenerated salt solution. *Environ. Sci. Technol.* **46,** 5240–5246 (2012).
81. Stoughton, R. W. & Lietzke, M. H. Calculation of some thermodynamic properties of sea salt solutions at elevated temperatures from data on NaCl solutions. *J. Chem. Eng. Data* **10,** 254–260 (1965).
82. Wick, G. L. & Isaacs, J. D. Salt domes: is there more energy available from their salt than from their oil? *Science* **199,** 1436–1437 (1978).
83. Loeb, S. Osmotic power plants. *Science* **189,** 654–655 (1975).

Acknowledgments The authors acknowledge support from the King Abdullah University of Science and Technology (KAUST) by Award KUS-I1-003-13 (BEL), and the World Class University (WCU) Program (Case III) through the National Research Foundation of Korea and the Ministry of Education, Science and Technology (R33-10046) (ME).

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/dujuno. Correspondence should be addressed to B.L. (blogan@psu.edu).

REVIEW

doi:10.1038/nature11478

Microbial engineering for the production of advanced biofuels

Pamela P. Peralta-Yahya[1,2], Fuzhong Zhang[1,2], Stephen B. del Cardayre[3] & Jay D. Keasling[1,2,4,5]

Advanced biofuels produced by microorganisms have similar properties to petroleum-based fuels, and can 'drop in' to the existing transportation infrastructure. However, producing these biofuels in yields high enough to be useful requires the engineering of the microorganism's metabolism. Such engineering is not based on just one specific feedstock or host organism. Data-driven and synthetic-biology approaches can be used to optimize both the host and pathways to maximize fuel production. Despite some success, challenges still need to be met to move advanced biofuels towards commercialization, and to compete with more conventional fuels.

Concerns about energy security, the global petroleum supply and climate change have increased interest in the production of sustainable and renewable liquid transport fuels[1,2]. The first-generation biofuels, ethanol and biodiesel, are the most widely used of these transport fuels[3]. The use of ethanol is due mainly to its high level of natural production by microbes, through fermentation of cornflour (in the United States) or cane sugar (in Brazil), and not because it is optimal for our existing petroleum-centric transport infrastructure. Ethanol has only 70% of the energy content of gasoline (petrol), it has a high tendency to absorb water from the air, which leads to corrosion in engines and pipes; and its distillation from the fermentation broth is energy intensive[1,4]. Biodiesel is the main biofuel in Europe, and is produced by the transesterification of vegetable oil or animal fats with methanol, but this also has its own limitations: it has only 91% of the energy content of D2 diesel and, because wax can form in the fuel if the temperature is too low, it is difficult to transport with the current distribution infrastructure, so there are geographical limits to its use.

Producing a high volume of first-generation biofuels from food crops creates a link between food and fuel prices. Instead of using starch and sucrose, biofuel production could come from more abundant and underused resources such as lignocellulose, algal biomass and greenhouse gases such as carbon monoxide and carbon dioxide. The non-food energy feedstock lignocellulose is the most abundant biomass on Earth, and consists of about 70% sugars. These sugars, however, require thermal, chemical and biochemical processes before they can be released for microbial fermentation. Algal biomass is a compelling alternative because it can be produced in salt water rather than on arable land, but its collection and dewatering is challenging. Atmospheric CO_2 can be transformed into biofuels through carbon fixation using engineered photosynthetic organisms. Carbon monoxide, from the thermal conversion of lignocellulosic biomass and abundant in steel-mill flu gas, can also be metabolized by microorganisms.

Progress in metabolic engineering, and synthetic and systems biology, have allowed the engineering of microbes to produce advanced biofuels with similar properties to petroleum-based fuels[5,6] (Table 1). When developing an organism or a pathway to produce an advanced fuel, factors including engine type (spark or compression ignition), energy content, combustion quality or ignition delay, cloud point, volatility, lubricity, viscosity, stability, odour, toxicity, water miscibility and cost must be considered[7,8]. The main challenge in using native hosts to convert feedstocks into advanced biofuels is to overcome the endogenous regulation of biofuel-producing pathways to achieve high yields. Reconstruction of advanced biofuel pathways in genetically tractable heterologous hosts, such as *Escherichia coli* and *Saccharomyces cerevisiae*, can work, but this approach presents its own challenges in balancing the enzyme activities and expression to maximize metabolic flux. Less genetically tractable hosts that have high biofuel tolerance or the ability to use non-sugar substrates are interesting alternatives because different feedstocks can be used. In this Review, we discuss the metabolic engineering of the pathways used in the production of advanced biofuels, irrespective of the feedstock and host organism, review data- and synthetic-biology-driven approaches for host and pathway optimization and discuss future directions in the microbial production of advanced biofuels (Fig. 1).

Pathways for alcohol-derived fuels

Short-chain or higher alcohols can be added to gasoline as oxygenates or, in some cases, to replace gasoline altogether. Butanol, for example, has 84% of the energy content of gasoline, limited miscibility with water and is completely miscible with gasoline[9]. In an effort to improve butanol production, species of the natural host *Clostridium* have been engineered to use feedstocks such as glucose[10], liquefied cornflour[11], glycerol (a by-product in the production of biodiesel from fats)[12] and even syngas (a mixture of hydrogen and carbon monoxide)[13]. However, butanol yields in this strict anaerobe seem to be limited by its slow growth rate and intolerance to butanol above 13 g l^{-1}, as well as our limited understanding of its physiology and genetics, and the paucity of synthetic-biology tools to engineer it[9,14,15].

Some of *Clostridium*'s limitations can be overcome by the introduction of its butanol pathway into organisms that grow faster, can tolerate high concentrations of butanol or can metabolize alternative feedstocks. *E. coli* has a high growth rate[16]; *S. cerevisiae* has a high tolerance to ethanol and potentially to butanol[17]; *Pseudomonas putida* can overcome toxicity using efflux pumps; *Bacillus subtilis* can change its cell-wall composition in response to solvent toxicity[18]; *Lactobacillus brevis* digests C_5 and C_6 substrates and has a high tolerance to butanol[19]; and the cyanobacterium *Synechococcus elongatus* is able to produce butanol from CO_2 by photosynthesis[20].

Regardless of the host organism, high-titre production of butanol is limited by the reversibility and cofactor specificity of the enzymes

[1]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, California 94608, USA. [2]QB3 Institute, University of California, Berkeley, California 94720. [3]LS9, 600 Gateway Boulevard, South San Francisco, California 94080, USA. [4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [5]Department of Chemical and Biomolecular Engineering, Department of Bioengineering, University of California, Berkeley, Berkeley, California 94720, USA.

in the *Clostridium* pathway (Fig. 2). However, this barrier to high butanol production was broken in *E. coli* by using a synthetic butanol pathway that combines enzymes from four different organisms chosen for their biochemical characteristics. Rather than using *Clostridium*'s reversible, flavin-dependent, oxygen-sensitive butyryl-CoA dehydrogenase complex, the synthetic pathway used biosynthetic NADH-dependent *trans*-enoyl-CoA reductase[21]. Increases in NADH, acetyl-CoA and ATP pools by deleting competing pathways and stripping the butanol from the reactor as it is produced have produced 30 g l^{-1} of butanol in fermenters[22]. The reversibility of the first step in the *Clostridium* butanol pathway has also been addressed. To favour the production of acetoacetyl-CoA (Fig. 2), instead of thiolase, acetyl-CoA decarboxylase driven by ATP and using CO_2 was used to generate malonyl-CoA and coupled to acetoacetyl-CoA synthase to produce acetoacetyl-CoA releasing CO_2 (ref. 20).

An alternative route for the production of alcohols, including butanol, is the Ehrlich, or 2-keto-acid, pathway. This pathway decarboxylates keto acids, the immediate amino-acid precursors, into aldehydes and reduces them to alcohols. The Ehrlich pathway is endogenous to some yeast species, in which the by-products of fermentation are fusel alcohols. Amino-acid-based alcohols, such as *n*-propanol from isoleucine, isobutanol from valine and *n*-butanol from norvaline, can be produced using this pathway[23]. Of these, isobutanol is the closest to industrial use, and although it has a similar energy content to *n*-butanol, its branching gives it improved properties, such as a better octane number (a measure of a fuel's resistance to knocking in spark ignition engines).

Introducing a promiscuous 2-keto-acid decarboxylase and an alcohol dehydrogenase into *E. coli* can produce isobutanol in high yields through the 2-keto-acid pathway[4]. Overexpression of 2-ketoisovalerate biosynthetic genes, deletion of multiple pathways competing for pyruvate consumption and the replacement of the endogenous acetolactate synthase enzyme with one from *B. subtilis* — which has a higher specificity for pyruvate — resulted in 22 g l^{-1} of isobutanol. The 2-keto-acid pathway has also been engineered in *Corynebacterium glutamicum*, a bacterium known for its high amino-acid production[24]; *S. cerevisiae*, in which the endogenous 2-ketoisovalerate biosynthetic pathway was upregulated and localized to the cytosol[25]; *Clostridium acetobutylicum*, a cellulolytic bacterium that produces isobutanol directly from crystalline cellulose through consolidated bioprocessing[26]; *S. elongatus*, a photosynthetic bacterium[27]; and *E. coli* engineered to metabolize protein-rich substrates such as the potentially abundant feedstock algal hydrolysate[28]. Isobutanol has also been produced using CO_2 obtained from electrochemically produced formate as the energy source for *Cupriavidus necator* (also known as *Ralstonia eutropha*)[29].

Pathways for isoprenoid-derived fuels

A class of compounds widely used as flavours and pharmaceuticals[30], isoprenoids have the potential to serve as advanced biofuels because of the branches and rings found in their hydrocarbon chain[7,8]. The branching means there are more tertiary carbon atoms to stabilize pressure-induced radicals, reducing premature ignition and increasing the octane number. The isoprenoid-based alcohol isopentanol is a potential substitute for gasoline. Although branching lowers the cetane number (a measure of combustion quality in compression engines) in diesels, having a few branches in an otherwise straight-chain alkane disrupts stacking and lowers the freezing point. The C_{15} isoprenoids farnesane and bisabolane have cetane numbers (of 58 and 52, respectively) that fall within the expected range for diesel (40–60). Jet-fuel replacements require high-energy content, which can be achieved by the presence of constrained ring structures, such as those found in pinene dimers[31].

Table 1 | Liquid transportation fuel properties and potential biofuels

Fuel type	Major components	Properties	Potential advanced biofuels
Gasoline	C_4–C_{12} hydrocarbons Linear, branched, cyclic aromatics	Octane number (87–91) Energy content	Butanol, isobutanol, short-chain alcohols, short branched-chain alkanes
Diesel	C_9–C_{23} hydrocarbons Linear, branched, cyclic aromatics	Cetane number (40–60) Good cold properties	Fatty alcohols, alkanes, linear or cyclic isoprenoids
Jet fuel	C_8–C_{16} hydrocarbons Linear, branched, cyclic aromatics	Heat density Very low freezing temperature	Branched alkanes, linear or cyclic isoprenoids

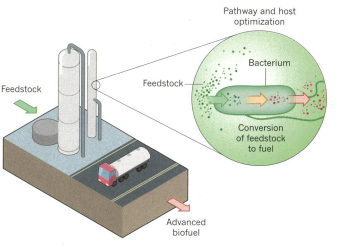


Figure 1 | The production of advanced biofuels. Engineering of bacteria or yeast species through data- or synthetic-biology-driven techniques optimizes the production of advanced biofuels, such as butanol, farnesane and bisabolane from feedstocks (ideally non-food resources such as lignocellulosic biomass, greenhouse gases and algal biomass).

Isoprenoid-based biofuels are produced from two universal C_5 precursors, isopentenyl diphosphate and dimethylallyl diphosphate[32], using either the mevalonate or deoxyxylulose-5-phosphate pathways (Fig. 3). Plants are natural sources of isoprenoids, but they do not produce large enough quantities to support biofuel production. Algae, such as *Botryococcus braunii*, produce large quantities, but they grow slowly and produce large amounts of fatty acids, rendering an extract akin to a biocrude, which would require cracking to form biofuels[33,34]. To increase the quantities produced, the deoxyxylulose-5-phosphate and mevalonate pathways (Fig. 3) can be deregulated or overexpressed in their native *E. coli* and *S. cerevisiae*, respectively, and introduced heterologously into these microorganisms[35–40]. Engineering of the mevalonate pathway resulted in 27 g l^{-1} of the C_{15} amorphadiene in *E. coli*[37] and 40 g l^{-1} in *S. cerevisiae*[38].

Of the isoprenoid-based biofuels, farnesane is the closest to commercialization[41,42]. The renewable products company Amyris based in Emeryville, California, uses the industrial yeast strain *S. cerevisiae* PE-2 to produce farnesene[43]. This is then chemically hydrogenated to farnesane, which is being evaluated as a high-performance advanced biofuel[41]. Bisabolane is another potential biofuel produced through a hybrid process, using microbial catalysis for bisabolene overproduction then chemical catalysis for reduction into bisabolane[44]. To produce bisabolene in *E. coli*, five plant isoprenoid synthases were screened for efficient conversion of farnesyl diphosphate into bisabolene, and the protein expression of the best one was improved using codon optimization for *E. coli*. The levels of the farnesyl diphosphate precursor were increased by codon-optimizing heterologous mevalonate pathway genes and by introducing a second promoter to upregulate transcription of bottleneck genes. Bisabolene was also produced using *S. cerevisiae*, and production by both organisms resulted in more than 900 mg l^{-1} (ref. 44). The discovery of bisabolene synthase as a flux-controlling enzyme, prompted the identification of the crystal structure of the most efficient bisabolene synthase[45] — from *Abies grandis* — to aid in its engineering for increased microbial bisabolene production.

Isopentenol and pinene can be chemically converted into isopentanol

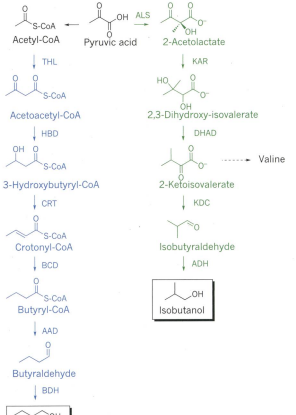


Figure 2 | Metabolic pathways used for the production of alcohol-based biofuels. In blue, *Clostridium*'s butanol pathway converts acetyl-CoA into butanol. AAD, butyraldehyde dehydrogenase; BCD, butyryl-CoA dehydrogenase; BDH, butanol dehydrogenase; CRT, crotonase; HBD, 3-hydroxybutyryl-CoA dehydrogenase; THL, thiolase. In green, the 2-keto acid pathway produces isobutanol from pyruvic acid. ADH, alcohol dehydrogenase; ALS, acetolactate synthase; DHAD, dihydroxy-acid dehydratase; KAR, ketol-acid reductoisomerase; KDC, keto-acid decarboxylase.

and pinene dimers[31]. Isopentenol was attained through the dephosphorylation of isopentenyl diphosphate using a phosphatase identified by screening a library of *B. subtilis* complementary DNA[46]. Pinene production in *E. coli* was achieved by re-routing isopentenyl diphosphate and dimethylallyl diphosphate using a geranyl diphosphate synthase and a pinene synthase (Fig. 3). The pinene-specific enzymes were introduced into an *E. coli* strain able to metabolize both xylose and cellulose in a co-culture to produce pinene at 1.5 mg l^{-1} from ionic-liquid-treated switchgrass as the sole carbon source[47].

Pathways for fatty-acid-derived fuels

Fatty acids in the form of phosphoglycerides and triglycerides are the main components of cell membranes, vegetable oils and animal fats, and their hydrophobic acyl chains can be used to produce biofuels. Technology to exploit fatty-acid metabolism is producing an expanded repertoire of fatty-acid-derived chemicals, many of which seem to be suitable alternatives to diesel.

Fatty acids are biosynthesized naturally by a large, multienzyme system called fatty-acid synthase (FAS), using malonyl-CoA as a building block (Fig. 4). Fatty acyl chains are elongated on acyl carrier proteins (ACPs) through repeated cycles of decarboxylative condensation, β-keto reduction, dehydration and enol reduction. Long-chain fatty acids can be released from the ACP by thioesterase-catalysed hydrolysis. The overexpression of a truncated, cytosolic thioesterase from bacteria[48] or plants[49,50] in *E. coli* produces fatty acids in the range of C_8 to C_{18}, depending on the thioesterase expressed. Hydrolysis of acyl-ACP is thought to relieve the feedback inhibition of FAS by the acyl-ACP[51], allowing efficient turnover of FAS and significant overproduction of fatty acids. In combination with the deletion of genes in the fatty-acid degradation pathway (β-oxidation), free fatty acids have been produced at high titres (1.2 g l^{-1}) (ref. 48), providing a large precursor pool for further conversion to advanced diesel fuel.

Because of the ionic nature of their carboxyl group, fatty acids cannot be used directly as biofuel, but they can be readily converted into non-ionic, hydrophobic molecules, such as fatty alcohols, fatty-acid alkyl esters, alkenes and alkanes. Fatty alcohols, a potential biofuel[7], have been biosynthesized at up to 60 mg l^{-1} by the expression of an acyl-CoA synthase to activate fatty acids to acyl-CoAs, followed by reduction by an acyl-CoA reductase through an aldehyde intermediate[48,52]. An acyl-ACP reductase that produces fatty aldehydes and alcohols directly when expressed in *E. coli*[53] has now been discovered, eliminating the need to go through the fatty-acid intermediate.

Methyl ketones in the C_{13} to C_{15} range have favourable cetane numbers and have been produced in engineered *E. coli* at 380 mg l^{-1} (ref. 54). In this pathway, the first three steps of β-oxidation are used to convert fatty acids into β-ketoacyl-CoAs, which are later hydrolysed by a thioesterase to form β-keto-fatty acids, followed by decarboxylation to yield methyl ketones. In addition, fatty-acid methyl and ethyl esters (FAMEs and FAEEs) — the main constituents of biodiesel used in trucks, buses and trains — have been biosynthesized at 1.5 g l^{-1} (ref. 55) by the activation of fatty acids to acyl-CoA, followed by ethanol esterification using a wax-ester synthase[48,56]. Alternatively, fatty acids can be methylated by *S*-adenosylmethionine to produce FAMEs[57] in one step using a fatty-acid methyltransferase. However, high-titre production of FAMEs from this pathway may be challenging because *S*-adenosylmethionine must be regenerated efficiently.

Alkanes are the main molecules in diesel fuel. Although some cyanobacteria naturally produce low levels of long-chain alkanes[53], the discovery of a biosynthetic pathway has allowed their heterologous production[58]. This has created an opportunity to produce long-chain alkanes directly from renewable materials in a single fermentation step and without the need for further chemical modification, such as hydrogenation. Alkane biosynthesis requires acyl-ACP reductase to catalyse the reduction of acyl-ACPs to form fatty aldehydes, and aldehyde decarbonylase to catalyse the decarbonylation of the aldehydes to formic acid and alkanes (or alkenes if the starting acyl groups are unsaturated). At least two other pathways exist for the biosynthesis of terminal alkenes. The first uses a cytochrome P450 enzyme that catalyses a decarboxylative oxidation to convert fatty acids to terminal alkenes[59]. The second involves a polyketide synthase, and produces terminal alkenes through a sulphonation-assisted decarboxylation[60]. A detailed description of the enzymatic mechanism has yet to be reported, and the pathway needs to be optimized to improve efficiency and yield. Finally, the production of long-chain (C_{24}–C_{31}) alkenes by a head-to-head condensation of two fatty acids followed by several reduction steps is also possible[61,62]. However, these long-chain hydrocarbons are waxes and would require further chemical processing, such as hydrocracking or reforming, before being used as fuels.

The building block of FAS-catalysed fatty-acid biosynthesis, malonyl-CoA, is converted from acetyl-CoA and bicarbonate by a multisubunit acetyl-CoA carboxylase. In prokaryotes and eukaryotes, this enzyme is tightly regulated at the transcriptional, translational[63] and post-translational[64] levels, leading to low cellular concentrations of malonyl-CoA. Furthermore, carboxylation of acetyl-CoA to malonyl-CoA consumes one ATP, and the incorporated CO_2 is returned to the solvent on decarboxylative Claisen condensation catalysed by 3-keto-acyl-ACP synthase. Biochemical Claisen condensations, such as those involved in the first step of butanol and polyhydroxyalkanoate biosynthesis, do not involve a decarboxylative mechanism. The consumption of energy and the regulation associated with acetyl-CoA carboxylase might be avoided

by synthesizing alkyl chains by running β-oxidation, or similar reactions, in reverse[65,66]. Reversing the β-oxidation cycle has been used to produce a diversity of C_4–C_{12} fatty species[67] by overexpressing several fatty-acid degradation enzymes in combination with an unconventional approach of global deregulation and strategic pathway knockouts. The results suggested high levels of butanol and fatty acids could be produced (roughly 7 g l^{-1} of fatty acids and about 14 g l^{-1} of *n*-butanol); however, which of the many deregulated enzymes are responsible for the products and how much product was derived from acetyl-CoA compared with malonyl-CoA remains unclear. If this approach is as useful as it seems, it could pave the way for anaerobic production of higher alkyl-chain products.

Pathways for polyketide-derived fuels

The isoprenoid and fatty-acid biosynthetic pathways produce a variety of regularly branched, cyclic and linear hydrocarbons, but the degree

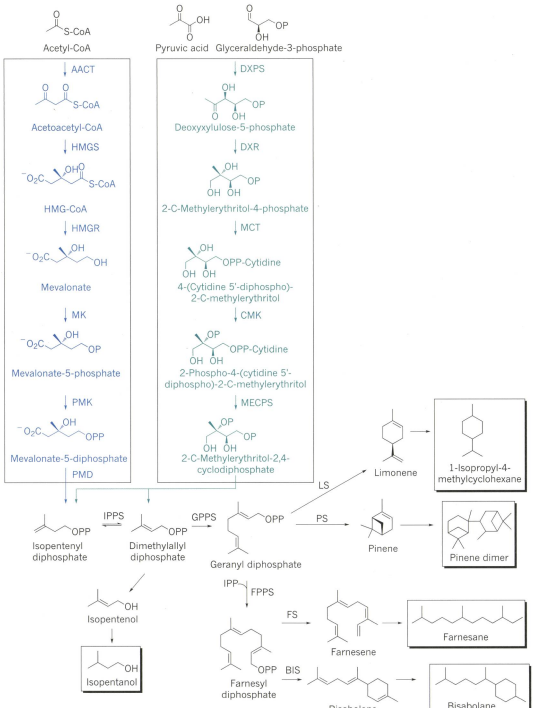


Figure 3 | Metabolic pathways used for the production of isoprenoid-based biofuels. In blue is the mevalonate pathway and in green is the deoxyxylulose-5-phosphate (DXP) pathway. AACT, acetyl-CoA transferase; BIS, bisabolene synthase; CMK, 4-(cytidine-5′-diphospho)-2-C-methylerythritol kinase; DXPS, DXP synthase; DXR, DXP reductoisomerase; FPPS, farnesyl diphosphate synthase; FS, farnesene synthase; GPPS, geranyl diphosphate synthase; HMGR, HMG-CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase; IPP, isopentenyl diphosphate; IPPS, IPP isomerase; LS, limonene synthase; MCT, 2-C-methylerythritol-4-phosphate cytidyltransferase; MECPS, 2-C-methylerythritol-2,4-cyclodiphosphate synthase; MK, mevalonate kinase; PMD, phosphomevalonate decarboxylase; PMK, phosphomevalonate kinase; PS, pinene synthase.

to which the hydrocarbons can be tailored is limited. The polyketide biosynthetic pathway is the most versatile for producing hydrocarbons with diverse structures, but it is relatively unexplored for fuels. This diversity means that they have been used as antibiotics, insecticides and antitumour agents[68]. Moreover, reduced polyketides are a good source of hydrocarbons that can produce biofuels as a derivative. They are biosynthesized through the decarboxylative condensation of malonyl-CoA by multidomain polyketide synthases, which share striking similarities with FASs. Polyketide synthases have been studied extensively over the past 30 years, creating an opportunity for engineering them to produce biofuels[69] (Fig. 4).

Modular type I polyketide synthases are the most extensively engineered[68,70], in part because they use different catalytic domains for each round of chain extension, providing the best control over polyketide structures. Each polyketide synthase contains several modules, a set of catalytic domains that are responsible for one round of polyketide chain extension and sub-sequence modification of the synthesized β-ketone. A downstream module can often process non-natural intermediates from the previous module, allowing the engineering of a modular type I polyketide synthase by several means: changing the precursor used for polyketide biosynthesis through the manipulation of the acyltransferase domains[71]; altering the structure of each extension unit by domain mutation, substitution and insertion within each module[72]; and changing the chain length of polyketides by module deletion, substitution or *de novo* design[70].

Optimization of hosts and pathways

High yields and productivities are crucial for the economical production of biofuels. Data-driven and synthetic-biology approaches for host and pathway optimization are showing promise in achieving these goals.

Data-driven approaches have the potential to streamline the engineering of microorganisms for the production of fuels and chemicals, possibly reducing the number of constructs needed to optimize a biofuel-producing organism. Functional-genomics approaches are well suited to identifying the rate-limiting steps in biosynthetic pathways, such as using targeted proteomics to identify two enzymes in the heterologous mevalonate pathway that were poorly expressed[73]. Increasing the enzyme expression levels through codon optimization of the relevant genes, in addition to increasing messenger RNA levels by the introduction of an additional 5′ promoter to the most poorly expressed gene, led to a three-fold improvement in sesquiterpene production[73]. By contrast, computer modelling of an organism's metabolism may be best suited to identifying off-pathway targets to increase biofuel production. For example, a metabolic model suggested that deletion of the NADPH-dependent glutamate dehydrogenase would lead to a ten-fold increase in the production of sesquiterpenes in *S. cerevisiae*[74]. Deleting the predicted gene, which increased the availability of NADPH for an important mevalonate pathway enzyme, nearly doubled sesquiterpene production[74]. Systems-biology approaches look beyond metabolism, identifying molecular mechanisms that could be applied to enhance biofuel production, and have been used to identify efflux pumps that

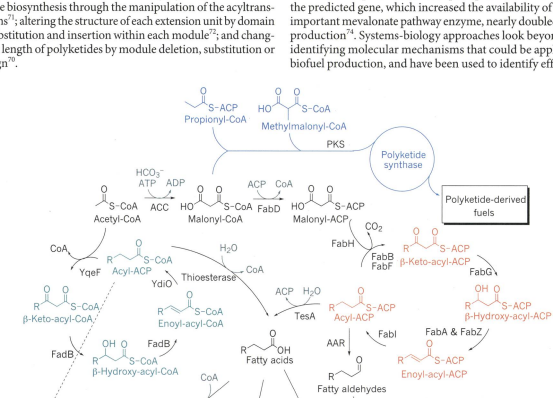


Figure 4 | Metabolic pathways used for the production of fatty-acid- and polyketide-derived biofuels. The fatty-acid biosynthetic cycle is in red, the reversal of the β-oxidation cycle is in green and polyketide synthase is in blue. AAR, acyl-ACP reductase; ACC, acetyl-CoA carboxylase; Acr1, acyl-CoA reductase; ADC, aldehyde decarbonylase; AtfA, wax-ester synthase; FabB, β-keto-acyl-ACP synthase I; FabD, malonyl-CoA:ACP transacylase; FabF, β-keto-acyl-ACP synthase II; FabG, β-keto-acyl-ACP reductase; FabH, β-keto-acyl-ACP synthase III; FabA & FabZ, β-hydroxyacyl-ACP dehydratase; FabI, enoyl-acyl-ACP reductase; FadB, enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; FadD, acyl-CoA synthase; FAMT, fatty acid methyltransferase; OleT$_{JE}$, *Jeotgalicoccus* sp terminal olefin-forming fatty-acid decarboxylase; PKS, polyketide synthase; TesA, acyl-ACP thioesterase; YdiO, enoyl-CoA reductase; YqeF, thiolase.

could aid in the production of biofuels that are toxic to the cell. Of the 43 potential biofuel efflux pumps that were identified from sequenced bacterial genomes, the expression of one pump improved *E. coli* survival in the presence of limonene, a potential biofuel precursor, and increased yield[75].

An extensively tested approach is to control flux of the biosynthetic pathways directly at the transcriptional, translational and post-translational levels (Fig. 5). Metabolic pathways can be controlled in several ways. The number of copies of a particular gene can be manipulated by increasing how many are integrated into the genome[76] or by placing the gene in a vector with high or low copy number[77]. The strength of constitutive and inducible promoters can be chosen to match the desired transcription initiation rate[78], and orthogonal, inducible promoters can be used to regulate parts of pathways, or multiple pathways independently[79]. Synthetic terminators can be engineered to control transcription termination efficiencies[80]. The metabolic pathway can be modulated at the translational level by inserting functional RNA segments into intergenic regions of operons to regulate the processing[81] and stability[82] of mRNAs that encode individual pathway enzymes. The strength of ribosome-binding sites can be predicted and designed using computational methods, allowing the control of the protein translation efficiencies[83]. Enzyme stability can be altered through the control of protein degradation rate using peptide tags[84].

These controls are effective in the engineered pathways, which are optimized on the basis of an initial set of conditions. However, the conditions may change when used in bioreactors and the control system cannot respond, potentially leading to suboptimal production of biofuels[85]. To address this, researchers have developed a dynamic sensor-regulator system (DSRS) that can adjust the metabolic pathway according to the metabolic status of the host and regulate the production of FAEE in engineered *E. coli*[55]. In this strategy, a biosensor that detects the cellular concentration of an intermediate acyl-CoA was used, and its cognate regulators — promoters responsive to the sensor — were engineered to control the expression of genes involved in FAEE biosynthesis. The DSRS regulated both the production and the consumption of pathway intermediates dynamically, according to the cellular concentration of acyl-CoA. As a result, the DSRS not only increased FAEE yield by threefold, but also reduced the cellular concentration of toxic intermediates and improved the genetic stability of FAEE-producing strains. Similar sensor-regulator systems that exert control at the translational level[86] can be developed to dynamically regulate other heterologous pathways.

Enzyme activity is a common bottleneck in the engineering of metabolic pathways. Many native enzymes are allosterically regulated by metabolites at the post-translational level; for example, *S. cerevisiae* mevalonate kinase is inhibited by several downstream intermediates, including farnesyl diphosphate. To overproduce farnesyl diphosphate, feedback-inhibition resistant mutants or homologues, such as mevalonate kinase from the archaeon *Methanosarcina mazei*[87], can be used to increase pathway flux. Furthermore, synthetic protein scaffolds could be used to spatially assemble metabolic enzymes and to catalyse multistep reactions synergistically[88]. When this system was applied to the heterologous mevalonate pathway in *E. coli*, mevalonate production was increased 77-fold, even though less protein was produced.

Moving to commercialization

To make the innovations in this Review sustainable, commercialization on a massive scale is needed. Advanced biofuels must be economically competitive with existing products, overcoming the primary economic drivers of feedstock price, and overall process productivity and yield. Although many of the products described have suitable performance characteristics, the potential yield for each product is limited by the theoretical yield of the production pathway[43], and this limit sets a bottom price that a product can achieve for a feedstock. Commercialization requires advancing laboratory-scale processes to yields and productivities that approach the theoretical (generally this will be 85% of what is theoretically possible) and scaling them up to reactors that will be more than 600 m^3 (roughly a million times larger than most lab-scale fermentations). The greatest challenges to commercialization are engineering catalysts to reach the yields and productivities required to meet economic targets, and scaling these processes without losing performance. Many of the advanced biofuels in the pipeline are stuck in an expensive purgatory, with only a few technologies for higher alcohols, butanol and isobutanol reaching the later phases of commercialization. These technologies are anaerobic processes that can be used in existing ethanol facilities, and have products that are naturally produced or engineered pathways linked to microbial fermentative growth[89,90]. Despite tremendous investments in isoprenoid and fatty-acid-derived products, commercial success has yet to be demonstrated. As these technologies become more refined, the more they differ with regard to industrial host, process and product trade-offs, and feedstock. Indeed, the closer commercialization becomes, the more crucial is the choice of host.

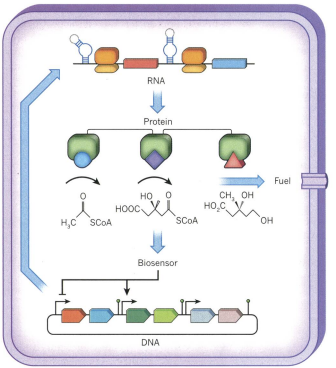


Figure 5 | Synthetic biology tools are used to maximize the production of biofuels from hosts. In DNA, genes involved in the synthesis of fuel can be introduced at the desired numbers into the host genome. Promoter (black arrows) and terminator sequences (green pins) can be engineered to control transcription rates. In the translation process, ribozymes (hairpins) can be inserted into functional RNA segments between the genes and ribosome binding sites can be engineered to control transcription initiation rates. Synthetic protein scaffolds incorporated into the pathway spatially assemble enzymes to efficiently transfer intermediates from one enzyme to the next in the pathway, increasing metabolic flux. Biosensors can detect specific metabolites in the host and regulate either transcription or translation dynamically, and the production of intermediates. Cellular pumps transport the fuel outside the cell (pumps).

Yeast's robustness, our extensive fermentation knowledge, the availability of genetic tools, its tolerance to industrial conditions and solvents (to butanol this is more than 20 $g\,l^{-1}$), its low medium pH and its lack of susceptibility to bacteriophage[91,92] make it the preferred organism for the production of butanol and isoprenoid-derived biofuels[91]. However, it is unable to digest the five-carbon sugars present in lignocellulosic biomass, such as xylose and arabinose; and its natural ability to produce ethanol can hinder metabolic engineering efforts to produce advanced biofuels. In addition, there are few synthetic biology tools for pathway optimization, and its lower protein expression levels compared with *E. coli* can limit the flux through biofuel producing pathways.

At least three renewable biofuel enterprises are pursuing

higher-alcohol production in yeast using slightly different strategies. Gevo based in Englewood, Colorado, has linked the production of isobutanol to anaerobic growth, and selected for strains that approach theoretical yields[89]. This, in combination with stripping the isobutanol from the broth through continuous flash evaporation, has resulted in more than 90% of the theoretical yield and a commercial-scale plant. Butamax in Wilmington, Delaware, has constructed various metabolic pathways that lead to butanol[90]. And Butalco in Fuerigen, Switzerland, has developed the technology to construct strains of yeast that can metabolize five-carbon sugars, and proposes using only endogenous genes to improve isobutanol production[84].

Exploitation of fatty-acid metabolism is being pursued in a variety of host organisms. Fatty-acid-derived compounds, such as FAMEs, fatty alcohols, alkanes and olefins, can be produced from *E. coli* in single-step fermentation from carbohydrate. LS9 in South San Francisco, California, chose *E. coli* to exploit because of its exceptionally high rate of fatty-acid biosynthesis (0.3 g l^{-1} per hour per gram of dry cell weight, based on the 30-minute doubling time and 9.7% lipid content of cell mass), its natural ability to secrete the compounds and consume both five-carbon and six-carbon sugars, the extensive industrial precedent in the commercial production of metabolically engineered small molecules (such as 1,3-propanediol, lysine and phenylalanine) and its ease of engineering. But, *E. coli* does have limitations, such as a preference for neutral pH and susceptibility to bacteriophages. Using cyanobacteria for the production of fatty-acid-derived compounds from CO_2 in photobioreactors is another option. The production of biodiesel and the renewable diesel precursor triacyl glyceride[85] have gravitated towards oleaginous algae; these GRAS (generally recognized as safe, classification by the US Food and Drug Administration) organisms naturally produce high levels of intracellular oil in both photobioreactors and heterotrophic fermentations[86]. However, until these technologies begin producing fuels at competitive prices, the jury is out on whether there will be widespread adoption.

Future directions

The past two years have seen advances in the use of different, inexpensive non-food feedstocks, and the creative engineering of metabolic pathways and enzymes for the production of advanced biofuels. The feasibility of these biofuel production pathways has been shown by the conversion of syngas, CO_2, algal hydrolysate and switchgrass into higher alcohols, fatty acids and isoprenoid-derived biofuels. Two years ago, the question would simply have been: which advanced biofuel can be made from glucose in high yield? But now researchers are also asking which feedstock can cost-effectively produce advanced biofuels in high yield. The challenge lies in producing a high yield of advanced biofuels, requiring the engineering of both the substrate use and advanced biofuel-producing pathways.

Advances in the production of alcohol-derived biofuels came from improving the kinetics of metabolic-pathway enzymes and generating metabolic driving forces to increase the flux through the pathways. Although biochemical characterization of metabolic-pathway enzymes is laborious, determining their biochemical parameters has allowed major breakthroughs in the production of biofuel. For example, *in vitro* reconstitution of the *E. coli* FAS determined the limits of the pathway for fatty-acid production[87]. For isoprenoid-derived biofuels, the effective union between biocatalysis and chemical catalysis has produced advanced biofuels, but the extensive in-pathway regulation by allosteric enzymes that respond to metabolic intermediates still needs to be overcome. Protein engineering could be used to remove or engineer the desired allostery into metabolic-pathway enzymes. Advances in chemical catalysis to allow the inexpensive conversion of biological compounds into fuels will help to make these a competitive alternative to conventional products.

Engineering the fatty-acid pathway has led to the production of several types of biofuel with different physical and combustion properties. Although the fuel properties of fatty alkanes are the most similar to conventional diesel fuel, the increased degree of unsaturation in alkenes may cause the formation of radicals during combustion. By contrast, oxygen-containing fatty methyl ketone and fatty-acid alkyl esters have higher cetane numbers and flashpoints, endowing shorter ignition delay and safer transport, respectively, but they have lower energy content and higher cloud points, which limits their use at very cold temperatures. Combining different types of fatty-acid-derived biofuel may allow fine-tuning for optimal fuel properties. Engineering the substrate specificity of FASs and the downstream biosynthetic pathways to alter length, functionality, and the degree of fatty-acid saturation and branching can further optimize fuel properties. In addition, engineered fuels with new features could be biosynthesized by polyketide synthases[88]. The challenges to this engineering will be designing the functional linkage between each domain and module, the proper folding and phosphopantetheinylation of the designed polyketide synthases[99], and the low catalytic efficiency and limited number of turnovers of engineered polyketide synthases. Advanced design tools and experimental techniques will facilitate the tuning of the polyketide-derived fuel properties by manipulating polyketide synthase.

Both data-driven and synthetic-biology approaches are powerful tools for trouble shooting and optimizing engineered metabolic pathways. Improving the efficiency and yield of biofuels further will require more potent tools, including techniques for stable maintenance of gene copy number and more precise control of mRNA and protein levels, especially in a dynamic manner that would automatically adjust the pathway according to its own metabolic flux. Furthermore, enzyme activities and substrate specificities can be altered by protein engineering[100]. Artificial enzymes with new functions can even be created by incorporating unnatural amino acids and computation-based protein design.

The development of pathways to expand the source of feedstock and to tune the properties of biofuel products, the creation of more metabolic engineering techniques to improve pathway flux and additional synthetic-biology approaches to optimize microbial hosts and boost production should see one or more advanced biofuels make it to market. ■

1. Stephanopoulos, G. Challenges in engineering microbes for biofuels production. *Science* **315,** 801–804 (2007).
2. Kerr, R. Global warming is changing the world. *Science* **316,** 188–190 (2007).
3. Fortman, J. L. *et al.* Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol.* **26,** 375–381 (2008).
4. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451,** 86–89 (2008).
This paper describes the rerouting of the *E. coli* amino-acid biosynthetic pathway for the overproduction of a variety of long-chain or higher alcohols.
5. Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* **330,** 1355–1358 (2010).
6. Nielsen, J. & Keasling, J. D. Synergies between synthetic biology and metabolic engineering. *Nature Biotechnol.* **29,** 693–695 (2011).
7. Lee, S. K. *et al.* Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* **19,** 556–563 (2008).
8. Peralta-Yahya, P. P. & Keasling, J. D. Advanced biofuel production in microbes. *Biotechnol. J.* **5,** 147–162 (2010).
9. Lee, S. Y. et al. Fermentative butanol production by *Clostridia*. *Biotechnol. Bioeng.* **101,** 209–228 (2008).
10. Qureshi, N. & Blaschek, H. P. Production of acetone butanol ethanol (ABE) by a hyper-producing mutant strain of *Clostridium beijerinckii* BA101 and recovery by pervaporation. *Biotechnol. Prog.* **15,** 594–602 (1999).
11. Ezeji, T. C., Qureshi, N. & Blaschek, H. P. Production of acetone butanol (AB) from liquefied corn starch, a commercial substrate, using *Clostridium beijerinckii* coupled with product recovery by gas stripping. *J. Ind. Microbiol. Biotechnol.* **34,** 771–777 (2007).
12. Malaviya, A., Jang, Y. S. & Lee, S. Y. Continuous butanol production with reduced byproducts formation from glycerol by a hyper producing mutant of *Clostridium pasteurianum*. *Appl. Microbiol. Biotechnol.* **93,** 1485–1494 (2011).
13. Kopke, M. *et al. Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc. Natl Acad. Sci. USA* **107,** 13087–13092 (2010).
14. Ezeji, T. *et al.* Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. *Appl. Microbiol. Biotechnol.* **85,** 1697–1712 (2009).
15. Lutke-Eversloh, T. & Bahl, H. Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. *Curr. Opin. Biotechnol.* **22,** 634–647 (2011).
16. Atsumi, S. *et al.* Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **10,** 305–311 (2008).
17. Steen, E. J. *et al.* Metabolic engineering of *Saccharomyces cerevisiae* for the

production of n-butanol. *Microb. Cell Fact.* **7,** 36 (2008).
18. Nielsen, D. R. *et al.* Engineering alternative butanol production platforms in heterologous bacteria. *Metab. Eng.* **11,** 262–273 (2009).
19. Berezina, O. V. *et al.* Reconstructing the clostridial *n*-butanol metabolic pathway in *Lactobacillus brevis*. *Appl. Microbiol. Biotechnol.* **87,** 635–646 (2010).
20. Lan, E. I. & Liao, J. C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl Acad. Sci. USA* **109,** 6018–6023 (2012).
21. Bond-Watts, B. B., Bellerose, R. J. & Chang, M. C. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nature Chem. Biol.* **7,** 222–227 (2011).
In this study, the authors assemble a synthetic butanol pathway in *E. coli* by choosing enzymes from different organisms on the basis of their biochemical properties.
22. Shen, C. R. *et al.* Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* **77,** 2905–2915 (2011).
23. Yan, Y. & Liao, J. C. Engineering metabolic systems for production of advanced fuels. *J. Ind. Microbiol. Biotechnol.* **36,** 471–479 (2009).
24. Smith, K. M., Cho, K. M. & Liao, J. C. Engineering *Corynebacterium glutamicum* for isobutanol production. *Appl. Microbiol. Biotechnol.* **87,** 1045–1055 (2010).
25. Chen, X. *et al.* Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. *Biotechnol. Biofuels* **4,** 21 (2011).
26. Higashide, W. *et al.* Metabolic engineering of *Clostridium cellulolyticum* for production of isobutanol from cellulose. *Appl. Environ. Microbiol.* **77,** 2727–2733 (2011).
27. Atsumi, S., Higashide, W. & Liao, J. C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nature Biotechnol.* **27,** 1177–1180 (2009).
28. Huo, Y. X. *et al.* Conversion of proteins into biofuels by engineering nitrogen flux. *Nature Biotechnol.* **29,** 346–351 (2011).
29. Li, H. *et al.* Integrated electromicrobial conversion of CO_2 to higher alcohols. *Science* **335,** 1596 (2012).
This paper presents the production of isobutanol in *Ralstonia* from CO_2 by using electricity.
30. Kirby, J. & Keasling, J. D. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu. Rev. Plant Biol.* **60,** 335–355 (2009).
31. Harvey, B. G., Wright, M. E. & Quintana, R. L. High-density renewable fuels based on the selective dimerization of pinenes. *Energy Fuels* **24,** 267–273 (2010).
32. Kuzuyama, T. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci. Biotechnol. Biochem.* **66,** 1619–1627 (2002).
33. Banerjee, A. *et al. Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals. *Crit. Rev. Biotechnol.* **22,** 245–279 (2002).
34. Tsukahara, K. *et al.* Treatment of liquid fraction separated from liquidized food waste in an upflow anaerobic sludge blanket reactor. *J. Biosci. Bioeng.* **87,** 554–556 (1999).
35. Martin, V. J. *et al.* Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnol.* **21,** 796–802 (2003).
36. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940–943 (2006).
37. Tsuruta, H. *et al.* High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLoS ONE* **4,** e4489 (2009).
38. Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl Acad. Sci. USA* **109,** E111–E118 (2012).
39. Leonard, E. *et al.* Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl Acad. Sci. USA* **107,** 13654–13659 (2010).
40. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330,** 70–74 (2010).
41. Renninger, N. & McPhee, D. Fuel compositions comprising farnesane and farnesane derivatives and method of making and using same. US patent 7,399,323 (2008).
42. Rude, M. A. & Schirmer, A. New microbial fuels: a biotech perspective. *Curr. Opin. Microbiol.* **12,** 274–281 (2009).
43. Ubersax, J. & Platt, D. Genetically modified microbes producing isoprenoids patent world 1,414,52 A1 (2010).
44. Peralta-Yahya, P. P. *et al.* Identification and microbial production of a terpene-based advanced biofuel. *Nature Commun.* **2,** 483 (2011).
In this paper, the authors identify bisabolane as a biosynthetic diesel and overproduce its immediate precursor, the isoprenoid bisabolene, in *E. coli* and *S. cerevisiae*.
45. McAndrew, R. P. *et al.* Structure of a three-domain sesquiterpene synthase: a prospective target for advanced biofuels production. *Structure* **19,** 1876–1884 (2011).
46. Withers, S. T. *et al.* Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. *Appl. Environ. Microbiol.* **73,** 6277–6283 (2007).
47. Bokinsky, G. *et al.* Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **108,** 19949–19954 (2011).
This paper presents the production of fatty-acid ethyl esters, butanol and the isoprenoid pinene in *E. coli* using ionic-liquid-treated switchgrass as the sole carbon source.
48. Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463,** 559–562 (2010).
This study describes several pathways leading to the bioproduction of fatty acids, fatty alcohols and fatty-acid ethyl esters in *E. coli* in high titres.
49. Lennen, R. M. *et al.* A process for microbial hydrocarbon synthesis: overproduction of fatty acids in *Escherichia coli* and catalytic conversion to alkanes. *Biotechnol. Bioeng.* **106,** 193–202 (2010).
50. Liu, T., Vora, H. & Khosla, C. Quantitative analysis and engineering of fatty-acid biosynthesis in *E. coli*. *Metab. Eng.* **12,** 378–386 (2010).
51. Handke, P., Lynch, S. A. & Gill, R. T. Application and engineering of fatty acid biosynthesis in *Escherichia coli* for advanced fuels and chemicals. *Metab. Eng.* **13,** 28–37 (2011).
52. Reiser, S. & Somerville, C. Isolation of mutants of *Acinetobacter calcoaceticus* deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *J. Bacteriol.* **179,** 2969–2975 (1997).
53. Tan, X. *et al.* Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. *Metab. Eng.* **13,** 169–176 (2011).
54. Goh, E. B. *et al.* Engineering of bacterial methyl ketone synthesis for biofuels. *Appl. Environ. Microbiol.* **78,** 70–80 (2011).
55. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnol.* **30,** 354–359 (2012).
In this study, the authors developed a dynamic regulatory system that allows the host cell to sense the concentration of a key intermediate and automatically regulate a heterologous pathway.
56. Kaltschemer, R., Stolting, T. & Steinbuchel, A. Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology* **152,** 2529–2536 (2006).
57. Nawabi, P. *et al.* Engineering *Escherichia coli* for biodiesel production utilizing a bacterial fatty acid methyltransferase. *Appl. Environ. Microbiol.* **77,** 8052–8061 (2011).
58. Schirmer, A. *et al.* Microbial biosynthesis of alkanes. *Science* **329,** 559–562 (2010).
This paper presents the discovery of two cyanobacterial genes responsible for the biosynthesis of alkanes.
59. Rude, M. A. *et al.* Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl. Environ. Microbiol.* **77,** 1718–1727 (2011).
60. Mendez-Perez, D., Begemann, M. B. & Pfleger, B. F. Modular synthase-encoding gene involved in alpha-olefin biosynthesis in *Synechococcus* sp. strain PCC 7002. *Appl. Environ. Microbiol.* **77,** 4264–4267 (2011).
61. Beller, H. R., Goh, E. B. & Keasling, J. D. Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Appl. Environ. Microbiol.* **76,** 1212–1223 (2009).
62. Sukovich, D. J. *et al.* Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Appl. Environ. Microbiol.* **76,** 3850–3862 (2010).
63. Meades, G. Jr *et al.* A tale of two functions: enzymatic activity and translational repression by carboxyltransferase. *Nucleic Acids Res.* **38,** 1217–1227 (2009).
64. Davis, M. S. & Cronan, J. E. Jr. Inhibition of *Escherichia coli* acetyl coenzyme A carboxylase by acyl-acyl carrier protein. *J. Bacteriol.* **183,** 1499–1503 (2001).
65. Seregina, T. A. *et al.* Construction of a butyrate-producing *E.coli* strain without the use of heterologous genes. *Appl. Biochem. Microbiol.* **46,** 745–754 (2010).
66. Sun, J., Pharkya, P. & Burgard, A. Primary alcohol producing organisms. US patent 0,040,426 A1 (2012).
67. Dellomonaco, C. *et al.* Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476,** 355–359 (2011).
This paper describes a unique pathway leading to the biosynthesis of alcohols and fatty-acid derivatives using reverse beta-oxidation cycles.
68. Yuzawa, S. *et al.* Heterologous production of polyketides by modular type I polyketide synthases in *Escherichia coli*. *Curr. Opin. Biotechnol.* Preprint at http://dx.doi.org/10.1016/j.copbio.2011.12.029 (2012).
69. Pfeifer, B. A. & Khosla, C. Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* **65,** 106–118 (2001).
70. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nature Biotechnol.* **23,** 1171–1176 (2005).
71. Jacobsen, J. R. *et al.* Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase. *Science* **277,** 367–369 (1997).
72. Yuzawa, S., Kim, L., Katz, L., *et al.* Heterologous production ofpolyketides by modular type I polyketide synthases in *Escherichia coli*. *Curr Opin Biotechnol.* http://dx.doi.org/10.1016/j.copbio.2011.12.029
73. Redding-Johanson, A. M. *et al.* Targeted proteomics for metabolic pathway optimization: application to terpene production. *Metab. Eng.* **13,** 194–203 (2011).
74. Asadollahi, M. A. *et al.* Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. *Metab. Eng.* **11,** 328–334 (2009).
75. Dunlop, M. J. *et al.* Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* **7,** 487 (2011).
76. Tyo, K. E., Ajikumar, P. K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature Biotechnol.* **27,** 760–765 (2009).
77. Jones, K. L., Kim, S. W. & Keasling, J. D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab. Eng.* **2,** 328–338 (2000).
78. De Mey, M. *et al.* Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol.* **7,** 34 (2007).
79. Lee, T. S. *et al.* BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *J. Biol. Eng.* **5,** 12 (2011).
80. d'Aubenton Carafa, Y., Brody, E. & Thermes, C. Prediction of rho-independent

Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.* **216,** 835–858 (1990).
81. Pfleger, B. F. *et al.* Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnol.* **24,** 1027–1032 (2006).
82. Smolke, C. D., Carrier, T. A. & Keasling, J. D. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Appl. Environ. Microbiol.* **66,** 5399–5405 (2000).
83. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnol.* **27,** 946–950 (2009).
84. Griffith, K. L. & Grossman, A. D. Inducible protein degradation in *Bacillus subtilis* using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. *Mol. Microbiol.* **70,** 1012–1025 (2008).
85. Holtz, W. J. & Keasling, J. D. Engineering static and dynamic control of synthetic pathways. *Cell* **140,** 19–23 (2010).
86. Michener, J. K. *et al.* Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab. Eng.* **14,** 212–222 (2011).
87. Primak, Y. A. *et al.* Characterization of a feedback-resistant mevalonate kinase from the archaeon *Methanosarcina mazei*. *Appl. Environ. Microbiol.* **77,** 7772–7778 (2011).
88. Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnol.* **27,** 753–759 (2009).
In this paper, a synthetic protein scaffold was used to spatially assemble metabolic-pathway enzymes to increase the flux through a heterologous metabolic pathway.
89. Bueller, T. *et al.* Engineered microorganisms capable of producing target compounds under anaerobic conditions. US patent 0,058,532 A1 (2012).
90. Donaldson, G. K. *et al.* Fermentative production of four carbon alcohols. US Patent 0,313,206 A1 (2007).
91. Hong, K. K. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell Mol. Life Sci.* http://dx.doi.org/10.1007/s00018-012-0945-1 (2012).
92. de Jong, B., Siewers, V. & Nielsen, J. Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. *Curr. Opin. Biotechnol.* http://dx.doi.org/10.1016/j.copbio.2011.11.021 (2011).
93. Donaldson, G. K. *et al.* Fermentative production of four carbon alcohols. US patent 0,313,206 A1 (2011).
94. Festel, G. *et al.* Fermentative production of isobutanol in yeast. US patent 0,053,235 A1 (2011).
95. Knothe, G. Biodiesel and renewable diesel: a comparison. *Prog. Energy Combust. Sci.* **36,** 364–373 (2010).
96. Trimbur, D. *et al.* Production of oil in microorganisms. US patent 0,047,863 A1 (2011).
97. Yu, X. *et al. In vitro* reconstitution and steady-state analysis of the fatty acid synthase from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **108,** 18643–18648 (2011).
98. Katz, L., Fortman, J. L. & Keasling, J. D. Producing biofuels using polyketide synthases. US patent CN102014623 (2011).
99. Betancor, L. *et al.* Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with *Streptomyces* chaperonins in *Escherichia coli*. *ChemBioChem* **9,** 2962–2966 (2008).
100. Tracewell, C. A. & Arnold, F. H. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* **13,** 3–9 (2009).

Acknowledgements P. P. P.-Y. and F. Z. contributed equally to this work. This work was supported in part by the Joint BioEnergy Institute, which is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231, and by the Synthetic Biology Engineering Research Center, which is funded by National Science Foundation Award No. 0540879.

Author Information Reprint and permission information is available at www.nature.com/reprints. The authors declare competing financial interests see go.nature.com/r3wogc for details. Readers are welcome to comment on this article at go.nature.com/r3wogc. Correspondence should be addressed to J.D.K. (keasling@berkeley.edu).

REVIEW

doi:10.1038/nature11479

Exploiting diversity and synthetic biology for the production of algal biofuels

D. Ryan Georgianna[1] & Stephen P. Mayfield[1]

Modern life is intimately linked to the availability of fossil fuels, which continue to meet the world's growing energy needs even though their use drives climate change, exhausts finite reserves and contributes to global political strife. Biofuels made from renewable resources could be a more sustainable alternative, particularly if sourced from organisms, such as algae, that can be farmed without using valuable arable land. Strain development and process engineering are needed to make algal biofuels practical and economically viable.

Despite limited supply and increasing demand, fossil fuels remain among the world's cheapest commodities. Prices will inevitably rise once demand starts to outstrip supply, but short- to medium-term replacement of fossil fuels by renewable and more environmentally benign alternatives will occur only if the substitutes can compete economically. One of these alternatives is based on the oils extracted from algae, and commercial-scale pilot facilities to test these are in operation. However, significant improvements are still needed to make algal biofuels economically viable. In this Review, we outline the advantages of algae as a biofuel producer, discuss the different cultivation methods, consider the options for achieving optimal algal biomass and lipid production, and the process engineering needed to make the process efficient and economically competitive.

Fuel from algae

Producing algal biofuel exploits the ability of algae to produce oils using only sunlight, carbon dioxide and water. Microalgae accumulate oil as nonpolar storage lipids, such as triacylglycerides (TAGs)[1]. The photosynthetic and cellular membranes of algae also contain polar lipids, such as glycolipids, phospholipids and sterols. Oils from algae can yield biodiesel through transesterification[2], and gasoline (petrol) or jet fuel through distillation and cracking[3].

Biofuel can be produced from various sources, but yield estimates are significantly higher for algae than for any other crop. This has considerable implications for land-area requirements: algae cultivated on only 30 million hectares and yielding biofuel at a conservative estimate of 40,000 litres per hectare per year is sufficient to replace the 1,200 billion litres of petroleum used by the world's largest consumer of petroleum, the United States (Fig. 1). The area is similar to that used for soya planting in the United States (about 29 million hectares) and roughly twice that used for the US production of corn ethanol (about 14 million hectares was used to produce almost 64 billion litres of ethanol in 2011), or an area about the size of New Mexico[4–6]. Furthermore, algal cultivation can use the large amount of non-arable land available for development without displacing food production, and its relatively high demand for water can be met by using low-quality sources such as waste or salt water[7].

Algae are almost ideal as organisms for developing the highly productive and robust crop strains that are essential for economically viable biofuel production. The search for these strains can exploit the vast diversity of algae, ranging from giant multicellular kelps to single-celled microalgae. Microalgae have been the focus of intense biofuel production efforts because of the ease, scale and speed at which they can be grown and manipulated, but strains differ significantly in lipid profile, photosynthetic ability, growth rate, growth medium requirement (from extreme halophilic to marine and fresh water), resistance to pathogens and biomass productivity. Although most algae are phototrophs, many can be grown heterotrophically.

Agricultural and industrial production

The underlying technologies and optimization strategies for producing biofuel through agricultural and industrial commercialization are very different.

In agricultural production, microalgae are cultivated in open ponds, with sunlight driving photosynthetic growth. This method uses non-arable land, consumes large amounts of CO_2 during the biomass production phase and, in principle, is extraordinarily scalable — limited only by space and capital costs. But growing algae efficiently and sustainably in fully exposed outdoor ponds is difficult, and suitable cultivation systems and practices are still under development.

Industrial production uses similar processes to industrial microbial fermentation in which yeast or bacteria are used as biorefining agents to produce food, beverages or high-value biotechnology products. For algae, production is driven by sunlight (in a photobioreactor) or reduced carbon sources such as sugar (in a fermentation reactor). Fermentation occurs in complete darkness and is identical to that of bacteria or yeast, which, as mature industrial technologies, have many of the systems and processes needed for algal fermentation already in place. However, fermentation has, so far, been used only to produce relatively high-value products; whether it can be developed for large-scale and economically viable fuel production is uncertain. In addition, algal fermentation of processed sugars to produce fuels heterotrophically does not produce lower greenhouse-gas emissions than the processing of fossil fuels[8], and is unlikely to equal their cost because the price of sugar is so high (more than US$0.66 per kilogram in 2011)[9]. Dependent on the development of advanced cellulose treatment technologies, using cellulosic biomass as the sugar source might be a solution[10]. However, this Review focuses on photobioreactors, which, like the agricultural process, rely on CO_2 and sunlight for photosynthetic production.

System optimization

Both agricultural and industrial production could be important in meeting the high demand for liquid fuels around the world. But their potential will be realized, and economic viability reached, only if the capital and energy costs of the production process can be greatly reduced and

[1]San Diego Center for Algae Biotechnology (SD-CAB), Division of Biological Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA.

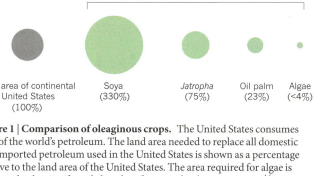


Figure 1 | Comparison of oleaginous crops. The United States consumes 25% of the world's petroleum. The land area needed to replace all domestic and imported petroleum used in the United States is shown as a percentage relative to the land area of the United States. The area required for algae is estimated to be significantly less than for any other biomass source[5,8].

the fuel yield enhanced.

The success of any biofuel production system will depend on economic and environmental costs. Life-cycle analysis (LCA) is a valuable tool for assessing this success, attempting to analyse the complete cost of a production process by considering the energy input and the environmental impact of each step. When combined with other economic analyses, LCA can help to identify high-cost areas, and determine whether a biofuel production strategy using a specific process and algal strain is viable. Also typically considered in an LCA, the basic processes involved in algal biofuel production, including the protection and maintenance of crops and the conversion of products into fuel (Fig. 2), are used to pinpoint where improvements will have a particularly strong affect on yield, environmental cost or energy use[11–13].

Not all LCAs take the same processes into account, which means the results can vary widely. For example, whether land-use change is considered[14] can be pivotal in determining whether a biofuel source reduces or increases greenhouse-gas emissions. Depending on the processes considered, annual yield estimates for agricultural algal biofuel production can vary between 8,000 l ha^{-1} and 140,000 l ha^{-1} (refs 15, 16). Overall output is of course important in determining the viability of a system, but even if the absolute numbers of LCA outputs vary, they still have a role in identifying the processes in need of optimization.

Unlike most agricultural systems, industrial production uses photobioreactors. These photobioreactors vary, but typically have a tubular or large, thin and flat panel design[17] (Fig. 3). Many factors can be adjusted to improve productivity and reduce costs, including culture media and nutrient sources, flow rates (amount of mixing) and sunlight exposure. Immobilized algal culture systems can be used for growth on a solid surface[17,18] to prevent shading and improve photosynthesis, which is especially useful for algae that secrete fuel precursors (Box 1), but early evaluation deemed this technology too expensive[19].

In a comparative LCA of the different growth systems, light-emitting diode (LED)-illuminated photobioreactors produced significantly more biomass (244,668 kg ha^{-1}) than solar-illuminated photobioreactors (8,262 kg ha^{-1}) and open ponds (4,957 kg ha^{-1}). But for the production costs per kilogram of biomass, this order was reversed: the estimated cost of production in open ponds was about $3 for a kilogram, five times lower than that achieved with solar-lit photobioreactors and eight times lower than with LED-lit photobioreactors[20]. So despite lower total biomass yield, the reduced input costs can result in a more efficient production system.

Irrespective of the organism used, the yield per input cost is the paramount factor for the control of overall economics. Improving algal growth while keeping input costs constant can neutralize many of the efficiency losses associated with the recovery of fuel. Yield is determined by the growth rate, cell density and lipid content of the algae, so optimization of each parameter affects this downstream processing efficiency. Higher growth rates reduce the time between harvesting, potentially reducing the exposure of algae to pathogens and the associated crop losses, and decreasing water consumption over a harvesting cycle. The final cell density of the culture — which determines the amount of biomass at harvest — is also important because harvesting is one of the more expensive processes. Engineering improvements will help to optimize efficiencies in biofuel processing, but the amount of biomass at harvest and its quality (lipid content) will affect productivity[21].

The yield of overall biomass and associated lipids that can be converted into fuel is controlled by a set of complex, and poorly understood, environmental and genetic factors. Achieving the best yield involves selecting promising strains (genotypes) and adjusting environmental factors, especially the growth-media composition. Although the relationship between metabolism and storage lipids is complex, understanding primary and secondary metabolism will allow strains to be engineered to suppress or remove processes that create energy-storing product molecules that are unsuitable for biofuel production. Conventional mutagenesis and genetic engineering have already been used to identify promising production strains, and recombinant DNA technologies are used to improve algal growth in both agricultural and industrial systems. Developing optimized production varieties will no doubt deliver profound yield improvements. Corn production today, for instance, has almost eight times the yield it did eight decades ago, before commercial varieties were developed[6].

Strain selection

Identifying the ideal algal strain for biofuel production is made easier by the immense diversity of the group, which has diverged over billions of years[22]. Algae have a potential genetic pool that is orders of magnitude larger than that of animals or land plants, but this diversity has only recently been examined. This potential is reflected in the diversity of algal species being explored for fuel production, which includes green algae, diatoms and cyanobacteria.

In agricultural systems, algal strains need to be able to tolerate variations in temperature, light, salinity and pathogen load. Strains optimized for cultivation in various geographical regions and in each season will be needed to ensure their widespread use. However, native strains are unlikely to have all of the characteristics needed for either agricultural or industrial production. No plant or animal used in large-scale agricultural or industrial production is the wild-type strain — they have all been domesticated by extensive modification through breeding and selection. A similar domestication process will be essential in obtaining the algal strains that generate adequate yield and quality of biomass.

Industrial bioreactors can be controlled, making them more stable environments than open ponds. Finding algal strains that grow quickly to high-cell densities and produce high levels of lipids using as little energy as possible will be essential to ensure cost-effective operations. Contamination is a significant factor in any cultivation system and, because stringent sterilization will almost certainly be too expensive, strains will need to have sufficient defence against or tolerance of pathogens to achieve high yields.

The temperature-control and gas-exchange systems needed in industrial systems to prevent alga death from excessive solar heating or oxygen poisoning are expensive components[23]. Most strains used in biofuel production grow best at 20–30 °C (ref. 23). Using thermophilic variants of cyanobacteria, which grow well between 35 °C and 50 °C, could be one solution to this expensive temperature control[24]. Solar heating is less of a concern with outdoor ponds, but lower growth during colder months — also applicable to outdoor photobioreactors — can significantly reduce yield. Locating ponds and outdoor photobioreactors in warm climates will compensate for these losses; however, ideally strains that produce high yields in a range of climates will be identified, or known strains will be modified. Many algae are already tolerant of cold temperatures and can recover from near freezing, but they grow slowly[25]. Several types of alga used for aquaculture are known to grow better at specific temperatures. For example, *Nannochloropsis* grows best at 25 °C; *Nannochloris* grows better at 30 °C; and *Chlorella vulgaris* has the best production at 10 °C (ref. 26). Rotation of algal strains based on growing season is therefore a likely production model, just as it is in terrestrial agriculture.

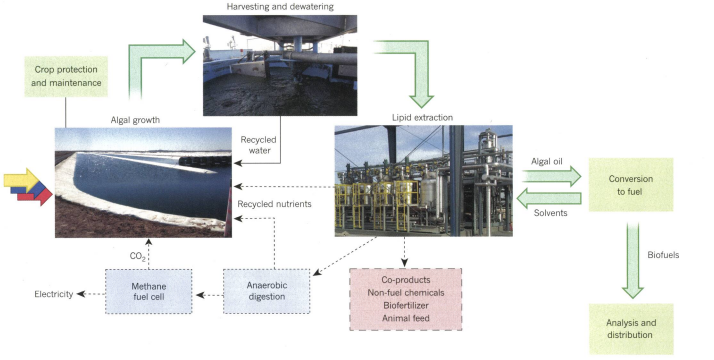


Figure 2 | Algal biofuel production. Light, water and nutrients (yellow, blue and red arrows) are required for algal growth in ponds. Some of the processes involved in algal biofuel production are common to most systems (green arrows). After fuel molecule extraction there are alternative uses for algal biomass (dashed arrows); many of these can produce co-products that are beneficial for economic and life-cycle analysis considerations. (Images courtesy of Sapphire Energy, San Diego, California).

A diverse set of wild-type strains have been isolated from different geographical and environmental sites, and evaluated for growth rate and storage-lipid accumulation[27–29]. Of 30 wild microalgal strains, *Nannochloropsis* sp. F&M-M26 was the most productive, achieving a lipid production of 61 mg l^{-1} per day through a biomass productivity of 0.21 g l^{-1} per day and a lipid content of 29.6% of the biomass[30]. The highest biomass productivity recorded is 0.37 g l^{-1} per day in *Porphyridium cruentum* (a red alga), and the highest lipid content is 39.8% in *Chaetoceros calcitrans* (a diatom); if these maximas were achievable in a single strain this could potentially be a 2.5-fold rise. Moving the most productive *Nannochloropsis* isolate to large photobioreactors increased lipid productivity to 84 mg l^{-1} per day. Despite this naturally occurring potential, the production efficiencies need to be much higher for commercial viability.

Strains of Chlorophyceae, Eustigmatophyceae, Haptophyceae and cyanobacteria are used to produce biofuels and other industrial organic chemicals[30], and many other types of algae have been tested. Although a natural isolate may exist that exceeds current yields, improving strains through breeding or recombinant DNA technologies is more likely to be successful in achieving the efficiencies needed for economic production.

Breeding to improve phenotypes

The main component missing from algal strain development is a breeding programme. Such programmes have been essential to the manipulation of plant crops over the past century[31], and continue to be used for yield improvement and to protect crops. The breeding systems are well described for the model green alga *Chlamydomonas*, but not for others. The reason may be, in part, a lack of mating compatible isolates — as interfertile isolates have been identified only for *Chlamydomonas reinhardtii* laboratory strains[32] — or the strong push in most biofuel research towards molecular and transgenic technologies rather than breeding. Strains of algae could be improved by searching for phenotypic heterogeneity within interfertile wild-type isolates and the possibility of transgression (extreme progeny phenotypes beyond those of either parent). Their short sexual cycles and rapid growth allow their quick selection from an extremely large genotypic pool, creating the potential for breeding in microalgae to exceed in a few years the gains obtained from a century of breeding programmes in conventional plant crops.

Genetic engineering

Engineering of algal metabolism has an important role in the improvement of growth and biomass accumulation. With the use of next-generation sequencing technologies, our understanding of algal genomes has improved rapidly, especially the genes involved in metabolic responses[33–35]. Analysis of nitrogen deprivation in algae, which increases the accumulation of TAGs, has identified many responsive genes, including 86 annotated as transcription factors[35]. Insertional mutagenesis of one of these factors can influence TAG accumulation[36], suggesting the potential for metabolic engineering for lipid accumulation.

Whether alga mutants grow faster than their wild-type progenitor is unknown, but improving the growth rates of biofuel-producing microalgae certainly has potential. Researchers have not been able to achieve efficient homologous recombination in the nuclear genome of the commonly transformed laboratory algal strain *C. reinhardtii*, which has limited some research, but this is not the case for the marine alga and biofuel candidate *Nannochloropsis*[37]. Although there is no known gene in *Nannochloropsis* to target to improve growth, an example from the yeast deletion collection shows what might be possible: a single gene deletion resulted in a 7% increase in growth rate[38]; if this growth occurred in algae, such an increase would double the cells and biomass obtained over the time it takes a wild-type strain to divide 15 times (about the normal harvest cycle for algae). Further annotation of algal genomes will assist in identifying potential genes to target for manipulating the growth rate.

Cell density depends on the media used and algal species involved, and is typically highest for heterotrophic growth in bioreactors. The chlorophyte *C. vulgaris* has a particularly high cell density during heterotrophic growth, reaching 120 g l^{-1} (ref. 39). At high densities, the secretion of molecules that mediate cellular communication becomes important. Both *C. vulgaris* and *C. reinhardtii* have been shown to

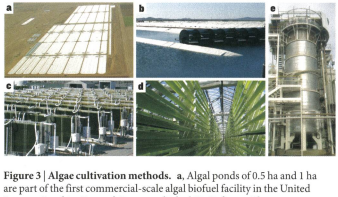


Figure 3 | Algae cultivation methods. **a**, Algal ponds of 0.5 ha and 1 ha are part of the first commercial-scale algal biofuel facility in the United States at Sapphire Energy's Integrated Algal BioRefinery. They cover an area 400 metres wide by 1,600 metres long at a location near Columbus, New Mexico. **b**, A single 1-million-litre paddle-wheel driven pond from the Columbus facility. **c**, A pilot-scale flat panel photobioreactor developed at the Laboratory for Algae Research and Biotechnology at Arizona State University in Mesa (image courtesy of Q. Hu). **d**, A commercial-scale tubular photobioreactor designed and constructed by IGV and operated by Salata in Germany (image courtesy of C. Grewe). **e**, An industrial-scale fermentation tank for heterotrophic cultivation of microalgae at Martek Biosciences, part of DSM in Heerlen, the Netherlands (image courtesy of D. Dong).

produce molecules that mimic quorum-sensing systems to inhibit bacterial growth[40,41]. Many other processes used to signal to other microalgal cells of the same species have also been identified, including autoinduction of cell division[42], premature cell death[43], allelopathy (production and release of metabolites that negatively affect the survival of another species[44]), and predation deterrence and warning systems for environmental stress[45,46]. Understanding cellular communication in algae might identify chemically regulated genes and allow the manipulation of pathways so that algal growth can exceed the current cell-density limits. For example, in *Escherichia coli*, a mutation in the *sdiA* gene, which is involved in the detection of *N*-acyl homoserine for quorum-sensing regulation, results in a 16% increase in final cell density compared with wild-type *E. coli*[47]. And in fungi, the molecule tyrosol can improve growth by selectively shortening the lag phase in a low-density culture without changing exponential growth[48].

Photosynthetic efficiency

The cultivation of algae in industrial photobioreactors or agricultural ponds aims to harvest as much solar energy as possible (Fig. 3). Efforts to improve photosynthetic efficiency have not been specific to algae; as a strategy, it has been proposed for increasing the yield of land plants to keep pace with increasing food demand where usable crop land is limited[49].

The genomes of most chloroplasts are capable of high rates of homologous recombination, and targeting specific genes in these organelles is not difficult in most species. Ideally, the photosynthetic characteristics of algae would be optimized for different photoperiods and light intensities. This optimization could be achieved by screening large numbers of mutants or by using the chloroplast photosynthetic machinery from species already adapted to these conditions. For instance, two strains of the green alga *Ostreococcus* with nearly identical morphology either grow near the surface (high-light strains) or at great depths (low-light strains), the antenna size in photosystems of low-light species is larger and, more strikingly, the electron flow is altered under low-light conditions[50]. Optimizing antenna size is also being considered in production systems as a potential means to limit the effect of shading and improve overall photosynthetic efficiency[51,52].

Manipulating the primary carbon-fixing enzyme Rubisco could also increase efficiency. Rubisco is the most abundant enzyme in all photosynthetic organisms, but also one of the least efficient. Hybrid enzymes from plant and algal species have already been shown to function in a different way from native enzymes[53]. In cyanobacteria, overexpression of Rubisco increased the photosynthetic rates (which is also potentially useful for the non-biomass derived biofuels in Box 1) when bicarbonate was used as a carbon source[54]. But Rubisco is not always rate limiting, its reduction (up to 50%) in plants and green algae did not have a negative effect on growth[55,56].

Photosystems within algae have evolved in response to their natural environments, and photosynthetic complexes are often remodelled in response to light, temperature or nutrient limitations[57]. Removing algae from their native environments protects them from occasional, but potentially lethal, environmental extremes. Engineering their chloroplasts to be optimal for the ideal environment, rather than the naturally occurring one, may improve photosynthetic efficiencies. A synthetic-biology approach has been used to introduce an exogenous chloroplast genome of *C. reinhardtii* into yeast cells[58]. A synthetic genome in yeast allows complex and rapid sequence manipulation to create synthetic chloroplast genomes; in fact, several photosystem proteins from *Scenedesmus obliquus* were found to function in *C. reinhardtii*[58]. This system could allow the creation of minimal plastids that are readily optimized for photosynthetic function.

Broadening lipid use

Lipid-productivity prospecting studies tend to examine neutral lipids (TAGs used for fatty-acid methyl esters). These constitute only a fraction of the total lipid content of an algal cell, which also includes polar lipids, such as phospholipids, sulpholipids and glycolipids. Using both types of lipids for fuel production would significantly increase the total yield of biodiesel that can be obtained from biomass. Academic and commercial enterprises have realized the importance of such research on this topic, and a direct transesterification method for total microalgal lipid content has produced appreciable levels of biodiesel even when there are undetectable levels of neutral lipids[59,60].

Genetic engineering could eventually enable algae to accumulate high levels of TAGs even under rapid growth conditions[61]. But being able to use essential polar lipids, as well as neutral ones, would ensure that a usable lipid source is available, irrespective of growth conditions.

Lipid and fuel quality

Lipid quality dictates many of the properties of the resultant fuel, including cetane number (a measurement of combustion quality in compression engines), oxidative stability, viscosity and cold flow[62]. Most common sources of biodiesel consist of unsaturated C_{16} and C_{18} fatty-acid methyl esters. Algae typically produce lipids between C_{14} and C_{20} in length. Unsaturated lipids usually yield fuels with a poor cetane number and low oxidative stability, whereas saturated fatty acids produce fuels at a higher cetane number and with greater stability but at the expense of their ability to be used at cold temperatures. Chemical strategies to improve the properties of biodiesel, such as reacting fatty-acid esters with alcohols other than methanol (ethyl and isopropyl esters have improved low-temperature properties), hydrogenation of lipids to increase saturation, traditional catalytic cracking and using additives to augment biofuel properties, can be used[62].

Attempts to increase lipid quantity through constitutive expression of transgenes in algae have had limited success owing to only minor expression of the transgene[63]. This is in contrast to overexpression in tobacco of the *Arabidopsis* genes *DGAT* (involved in TAG biosynthesis) and *LEC2* (a seed maturation and oil-storage regulatory gene), which increased the extracted fatty-acid yield[64]. However, lipid production is complex and may differ between some algae and higher plants, for example DGAT has been overexpressed in *C. reinhardtii* with no observable effect on lipid accumulation[65].

Another class of hydrocarbons found in plants and algae, the isoprenoids, are complex organic compounds that are ideal fuel precursors and could enhance the fungibility of biofuels[66]. Many isoprenoids are produced in the chloroplast, and a better understanding of their regulation and how best to modify the biosynthetic pathways might allow more biofuel precursors to be produced[67]. However, increasing lipid content and altering its type could have undesirable side effects: the green microalga

Botryococcus accumulates nearly 40% of its biomass in the form of the terpenoid hydrocarbon botryococcene[68] but at the expense of growth rate[69].

Nutrient use and recycling

Photosynthesis eliminates the need for costly organic carbon sources. However, media that can support a rich algal growth still require nutrients in the form of nitrogen, phosphorus and potassium, which are expensive[70] and — in the case of phosphorus — limited[71]. The cost can be managed by locating algal ponds or bioreactors near nutrient-rich wastewater streams, or by using feed sources such as anaerobic digester waste effluents[72]. Another option is to recycle the nutrients that remain in the algal waste after lipid extraction, rather than producing pellets to sell as a coproduct for animal feed. Evidence for the anaerobic digestion of algal waste[73,74] suggests that the next logical step is to optimize algae to grow on these waste streams. This strategy creates a process waste stream and provides nutrients, while using the energy stored within the waste algal biomass could help to run the production process. Chemical and engineering hurdles need to be overcome before this process will be feasible, and algae must be engineered to process these nutrients, including complex nitrogen and phosphorus, from the recycled algal waste. Other approaches for recycling these nutrients include supercritical water extraction to remove nearly all fatty acids but leave free ammonia and phosphates in an aqueous phase, which can then be used to supplement fresh growth media[75,76].

Glycerol recovery is also possible. This is one of the main by-products of the transesterification process that can serve as a carbon feedstock for heterotrophic algal growth. Grown on crude glycerol, *Chlorella protothecoides* has delivered biomass and lipid yields similar to growth on glucose[77]. Although the energy available in waste glycerol as the sole carbon source for energy production would not be viable, using this waste stream in photobioreactors as a feedstock supplement could certainly increase biomass yields.

Protection from pathogens

Crop protection is essential for outdoor pond systems and closed bioreactors. However, relatively little is known about the threats to algal farming even from known pathogens and grazing pests, and others may emerge with more intense farming. Long-term pilot outdoor facilities will be essential to determine the pathogen load in various environments. Pathogens of algae are extraordinarily diverse and include viruses, bacteria, fungi and other eukaryotes[78]. Apart from a virus infection of *Chlorella* discovered in the early 1980s (ref. 79), very few algal viruses have been studied. Grazing by zooplankton is also a pressure, but parasites are the greater threat to algal blooms in natural environments[80].

Strain selection can also greatly affect crop protection. Marine algae, such as *Dunaliella*, grow in halophilic media that few organisms can survive. *Nannochloropsis* species tolerate alkaline pH fluctuations[81], which may not be tolerated by many pathogens. Alkaliphilic cyanobacteria are already produced in large outdoor ponds for nutraceutical production in cultures that are almost entirely free of contamination.

As for many traditional crop plants, it may be possible to identify or breed algae that are resistant to the pesticides and antibiotics used to control unwanted species. Genetic transformation of algae has been established in many species aside from the model organism *Chlamydomonas*, and many of the methods used should be transferable to biofuel strains. Along with the requisite resistance gene, a low-cost and high-efficacy pesticide or antibiotic would need to be identified for treatment of ponds and bioreactors. Another, very different strategy for fighting bacterial contamination exploits the quorum-sensing systems that inhibit bacterial growth, and is currently being explored in mass cultivation of microalgae[82].

Improved harvesting

Harvesting, especially dewatering algal biomass, is one of the main costs in algal biofuel production. Most open ponds use a semicontinuous harvest strategy that removes roughly half of the algae, alleviating the need for re-inoculation. The algae are harvested after each doubling, so growth of a single generation per day would equate to a single harvest per day[83]. Alternatively, continuous systems to keep algae at a constant density and batch or complete systems are reminiscent of agricultural crops, removing nearly all algae during a single harvest.

Harvesting typically involves separation of the algae by sedimentation or flocculation, followed by filtration or centrifugation to remove the unwanted water. Drying the algae is an enormously energy-intensive step, which is one of the constraints on commercial development. Solar drying is an option, but still problematic with the current technology, owing to the time and space required.

Optimizing algae to ease harvesting can include increasing cell size, which helps with crop protection strategies and aids filtration. Algal strains can also be engineered such that the addition of a polymer or a change in an environmental variable triggers flocculation[84,85]. Conceivably, proteins involved in flocculation could be introduced and manipulated to respond to a particular growth state, as demonstrated in industrial strains of yeast[85,86]. Extraction of lipids from wet algae is possible using supercritical water extraction of algae with a moisture content of around 80% (ref. 76). Efficient wet extraction could remove the need for extensive dewatering and drying.

Outlook

Diminishing sources of fossil fuels and concerns over their use have made the need for efficient and sustainable biofuel production strategies more urgent. Large-scale biofuel production exists, but there is unease about its affect on food production and changes to land use. Algal biofuels could alleviate these worries. Algae can be grown quickly on non-arable land to produce large quantities of lipids that can be readily converted to fungible fuels. Although biofuel production from algae is still too expensive to compete with fossil fuels, it is an exciting prospective sustainable biofuel source.

BOX 1

Platforms for production

The designs for cultivation of algae to produce biofuel in ponds, bioreactors and photobioreactors all have unique characteristics, and how they produce fuel varies. In the agricultural model, different strains have been tested with different water sources and pond designs but, in all large-scale facilities, lipids are first accumulated inside the alga cell, then extracted after harvest and converted to fuel. The industrial model usually uses the same approach, whether using heterotrophic growth in fermenters or photosynthetic growth in photobioreactors.

Photobioreactors are expensive to scale, so direct secretion of fuel molecules from alga cells has attracted interest as a way of reducing the cost of harvesting. Fuels can be lipids that are naturally secreted by some algae[68,69] or algae engineered to secrete lipids from direct photosynthetic conversion[54,87]. Photofermentative growth is not a new concept for algae, and genetic modification has been aimed at lipid secretion and hydrogen production[88]. As algae are capable of accumulating sugars and starches from photosynthesis, conversion of these carbohydrates into ethanol, and secretion of this primary fuel has also been pursued[89]. In many cases, the accumulation of fuel and fuel precursors in an oxygenated environment will promote the proliferation of contaminating organisms, making crop protection strategies paramount to prevent catabolism of the secreted products. Although early-development fuel-secretion technology may reduce the need to harvest algae, any dense cultures used will eventually die, so the need to remove biomass from the growth medium is not completely eliminated (unless water and nutrient recycling are ignored).

With concerted effort, improvements can be made to the agricultural and industrial cultivation of algae for biofuel production. Industrial cultivation produces the most algae per litre, but is very expensive. Agricultural cultivation could in principle achieve cost efficiencies closer to those needed to compete with petroleum-based fuels, and with a reduced overall environmental impact, but its viability at a sufficiently large scale is only just beginning to be tested. Efforts to increase productivity but reduce input and cost through process engineering and the use of transgenic methods, and classical breeding aimed at developing domesticated algal crop strains, will benefit both strategies. With time, and driven by the continuing rise in the price of fossil fuels, technology and algal phenotypes will be developed to allow algal biofuel to compete with the commodities that currently underpin modern life — fossil fuels. ■

1. Guschina, I. A. & Harwood, J. L. Lipids and lipid metabolism in eukaryotic algae. *Prog. Lipid Res.* **45,** 160–186 (2006).
2. Amin, S. Review on biofuel oil and gas production processes from microalgae. *Energy Convers. Manage.* **50,** 1834–1840 (2009).
3. Luo, Y. *et al.* The thermal cracking of soybean/canola oils and their methyl esters. *Fuel Process. Technol.* **91,** 613–617 (2010).
4. US DOE. *National Algal Biofuels Technology Roadmap.* (United States Department of Energy, 2010).
5. USDA-NASS. http://www.nass.usda.gov (2012).
6. US Energy Information Administration. http://www.eia.gov (US Energy Information Administration, 2012).
7. Wigmosta, M. S., Coleman, A. M., Skaggs, R. J., Huesemann, M. H. & Lane, L. J. National microalgae biofuel production potential and resource demand. *Wat. Resour. Res.* **47,** http://dx.doi.org/10.1029/2010WR009966 (2011).
8. Darzins, A., Pienkos, P. T. & Edye, L. *Current status and potential for algal biofuels production.* IEA Bioenergy Task 39 (2010).
9. NYSE Euronext. *London International Financial Futures and Options Exchange* http://www.euronext.com (NYSE Euronext, 2012).
10. US Department of Energy. *Solazyme Integrated Biorefinery: Diesel fuels from heterotrophic algae.* <http://www.eere.energy.gov> (US Department of Energy, 2010).
11. Chowdhury, R., Viamajala, S. & Gerlach, R. Reduction of environmental and energy footprint of microalgal biodiesel production through material and energy integration. *Bioresour. Technol.* **108,** 108–111 (2012).
12. Khoo, H. H. *et al.* Life cycle energy and CO_2 analysis of microalgae-to-biodiesel: preliminary results and comparisons. *Bioresour. Technol.* **102,** 5800–5807 (2011).
13. Shirvani, T., Yan, X. Y., Inderwildi, O. R., Edwards, P. P. & King, D. A. Life cycle energy and greenhouse gas analysis for algae-derived biodiesel. *Energy Environ. Sci.* **4,** 3773–3778 (2011).
14. Searchinger, T. *et al.* Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **319,** 1238–1240, (2008).
15. Savage, N. Algae: the scum solution. *Nature* **474,** S15–S16 (2011).
16. Scott, S. A. *et al.* Biodiesel from algae: challenges and prospects. *Curr. Opin. Biotechnol.* **21,** 277–286 (2010).
17. Shen, Y., Yuan, W., Pei, Z. J., Wu, Q. & Mao, E. Microalgae mass production methods. *Trans. ASABE* **52,** 1275–1287 (2009).
18. Leonard, A. *et al.* Whole-cell based hybrid materials for green energy production, environmental remediation and smart cell-therapy. *Chem. Soc. Rev.* **40,** 860–885 (2011).
19. Jaycor. *Review and Evaluation of Immobilized Algae Systems for the Production of Fuels from Microalgae: a Final Subcontract Report.* (Solar Energy Research Institute, 1985).
20. Amer, L., Adhikari, B. & Pellegrino, J. Technoeconomic analysis of five microalgae-to-biofuels processes of varying complexity. *Bioresour. Technol.* **102,** 9350–9359 (2011).
This article presents an LCA, concluding that, despite higher biomass accumulation in photobioreactors, open ponds are more cost effective because of their significantly lower construction costs.
21. Davis, R., Aden, A. & Pienkos, P. T. Techno-economic analysis of autotrophic microalgae for fuel production. *Appl. Energy* **88,** 3524–3531 (2011).
22. Tirichine, L. & Bowler, C. Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *Plant J.* **66,** 45–57 (2011).
This article reviews genomic sequencing efforts in algae and describes diversity from a functional perspective.
23. Bechet, Q., Shilton, A., Fringer, O. B., Munoz, R. & Guieysse, B. Mechanistic modeling of broth temperature in outdoor photobioreactors. *Environ. Sci. Technol.* **44,** 2197–2203 (2010).
24. Miyamoto, K., Hallenbeck, P. C. & Benemann, J. R. Nitrogen-fixation by thermophilic blue-green-algae (cyanobacteria) — temperature characteristics and potential use in biophotolysis. *Appl. Environ. Microbiol.* **37,** 454–458 (1979).
25. Butterwick, C., Heaney, S. I. & Talling, J. F. Diversity in the influence of temperature on the growth rates of freshwater algae, and its ecological relevance. *Freshwat. Biol.* **50,** 291–300 (2005).
26. Bae, J. H. & Hur, S. B. Selection of suitable species of *Chlorella, Nannochloris,* and *Nannochloropsis* in high- and low-temperature seasons for mass culture of the rotifer *Brachionus plicatilis. Fish. Aquat. Sci.* **14,** 323–332 (2011).
27. Griffiths, M. J. & Harrison, S. T. L. Lipid productivity as a key characteristic for choosing algal species for biodiesel production. *J. Appl. Phycol.* **21,** 493–507 (2009).
28. Mutanda, T. *et al.* Bioprospecting for hyper-lipid producing microalgal strains for sustainable biofuel production. *Bioresour. Technol.* **102,** 57–70 (2011).
29. Rodolfi, L. *et al.* Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol. Bioeng.* **102,** 100–112 (2009).
This bioprospecting study examines 30 strains in a laboratory and 4 of these in large outdoor photobioreactors. A promising marine *Nannochloropsis* species that grows well under various conditions and accumulates lipids at high levels was identified.
30. Larkum, A. W., Ross, I. L., Kruse, O. & Hankamer, B. Selection, breeding and engineering of microalgae for bioenergy and biofuel production. *Trends Biotechnol.* **30,** 198–205 (2011).
31. Moose, S. P. & Mumm, R. H. Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* **147,** 969–977 (2008).
32. Coleman, A. W. & Mai, J. C. Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting genetic relatedness. *J. Mol. Evol.* **45,** 168–177 (1997).
33. Castruita, M. *et al.* Systems biology approach in *Chlamydomonas* reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell* **23,** 1273–1292 (2011).
34. Gonzalez-Ballester, D. *et al.* RNA-seq analysis of sulfur-deprived *Chlamydomonas* cells reveals aspects of acclimation critical for cell survival. *Plant Cell* **22,** 2058–2084 (2010).
35. Miller, R. *et al.* Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism. *Plant Physiol.* **154,** 1737–1752 (2010).
36. Boyle, N. R. *et al.* Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas. J. Biol. Chem.* **287,** 15811–15825 (2012).
The authors of this study use a large genome insertional library to identify mutants in certain genes, finding several that show altered lipid accumulation.
37. Kilian, O., Benemann, C. S., Niyogi, K. K. & Vick, B. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc. Natl Acad. Sci. USA* **108,** 21265–21269 (2011).
This article reports high rates of homologous recombination in a biofuel candidate eukaryotic algae.
38. Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nature Methods* **5,** 711–718 (2008).
39. Doucha, J. & Livansky, K. Production of high-density *Chlorella* culture grown in fermenters. *J. Appl. Phycol.* **24,** 35–43 (2012).
40. Rajamani, S. *et al.* The vitamin riboflavin and its derivative lumichrome activate the LasR bacterial quorum-sensing receptor. *Mol. Plant Microbe Interact.* **21,** 1184–1192 (2008).
41. Teplitski, M. *et al. Chlamydomonas reinhardtii* secretes compounds that mimic bacterial signals and interfere with quorum sensing regulation in bacteria. *Plant Physiol.* **134,** 137–146 (2004).
42. Grabski, K. & Tukaj, Z. Autoinduction activity of a conditioned medium obtained from high density cultures of the green alga *Scenedesmus subspicatus. J. Appl. Phycol.* **20,** 323–330 (2008).
43. Vardi, A. *et al.* Synchronization of cell death in a dinoflagellate population is mediated by an excreted thiol protease. *Environ. Microbiol.* **9,** 360–369 (2007).
44. Legrand, C., Rengefors, K., Fistarol, G. O. & Graneli, E. Allelopathy in phytoplankton — biochemical, ecological and evolutionary aspects. *Phycologia* **42,** 406–419 (2003).
45. Pohnert, G., Steinke, M. & Tollrian, R. Chemical cues, defence metabolites and the shaping of pelagic interspecific interactions. *Trends Ecol. Evol.* **22,** 198–204 (2007).
46. Vardi, A. *et al.* A stress surveillance system based on calcium and nitric oxide in marine diatoms. *PLoS Biol.* **4,** 411–419 (2006).
47. Lee, J., Maeda, T., Hong, S. H. & Wood, T. K. Reconfiguring the quorum-sensing regulator SdiA of *Escherichia coli* to control biofilm formation via indole and N-acylhomoserine lactones. *Appl. Environ. Microbiol.* **75,** 1703–1716 (2009).
48. Chen, H., Fujita, M., Feng, Q. H., Clardy, J. & Fink, G. R. Tyrosol is a quorum-sensing molecule in *Candida albicans. Proc. Natl Acad. Sci. USA* **101,** 5048–5052 (2004).
49. Zhu, X. G., Long, S. P. & Ort, D. R. Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* **61,** 235–261 (2010).
50. Cardol, P. *et al.* An original adaptation of photosynthesis in the marine green alga *Ostreococcus. Proc. Natl Acad. Sci. USA* **105,** 7881–7886 (2008).
51. Ort, D. R., Zhu, X. G. & Melis, A. Optimizing antenna size to maximize photosynthetic efficiency. *Plant Physiol.* **155,** 79–85 (2011).
52. Beckmann, J. *et al.* Improvement of light to biomass conversion by de-regulation of light-harvesting protein translation in *Chlamydomonas reinhardtii. J. Biotechnol.* **142,** 70–77 (2009).
53. Genkov, T., Meyer, M., Griffiths, H. & Spreitzer, R. J. Functional hybrid rubisco enzymes with plant small subunits and algal large subunits engineered *rbcS* cDNA for expression in *Chlamydomonas. J. Biol. Chem.* **285,** 19833–19841 (2010).
54. Atsumi, S., Higashide, W. & Liao, J. C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nature Biotechnol.* **27,** 1177–1142 (2009).
This article reports the engineering of the cyanobacterium *Synechococcus elongatus* to convert CO_2 into the secreted fuel precursor isobutyraldehyde at

high levels through the addition of several bacterial genes (*kivd*, *alsS*, *ilvC and ilvD*) and reports an increase in production efficiency through improvement of photosynthesis by overexpression of Rubisco.

55. Stitt, M. & Schulze, D. Does rubisco control the rate of photosynthesis and plant-growth — an exercise in molecular ecophysiology. *Plant Cell Environ.* **17,** 465–487 (1994).
56. Johnson, X. Manipulating RuBisCO accumulation in the green alga, *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* **76,** 397–405 (2011).
57. Moseley, J. L. *et al.* Adaptation to Fe-deficiency requires remodeling of the photosynthetic apparatus. *EMBO J.* **21,** 6709–6720 (2002).
58. O'Neill, B. M. *et al.* An exogenous chloroplast genome for complex sequence manipulation in algae. *Nucleic Acids Res.* **40,** 2782–2792 (2011).
In this article, the authors develop a system for maintaining and quickly manipulating an algal chloroplast genome in *S. cerivisiae* by reintroducing a synthetic chloroplast genome into algae.
59. Wahlen, B. D., Willis, R. M. & Seefeldt, L. C. Biodiesel production by simultaneous extraction and conversion of total lipids from microalgae, cyanobacteria, and wild mixed-cultures. *Bioresour. Technol.* **102,** 2724–2730 (2011).
60. Du, W., Xu, Y. Y., Zeng, J. & Liu, D. H. Novozym 435-catalysed transesterification of crude soya bean oils for biodiesel production in a solvent-free medium. *Biotechnol. Appl. Biochem.* **40,** 187–190 (2004).
61. Wijffels, R. H. & Barbosa, M. J. An outlook on microalgal biofuels. *Science* **329,** 796–799 (2010).
62. Knothe, G. Improving biodiesel fuel properties by modifying fatty ester composition. *Energy Environ. Sci.* **2,** 759–766 (2009).
63. Radakovits, R., Jinkerson, R. E., Darzins, A. & Posewitz, M. C. Genetic engineering of algae for enhanced biofuel production. *Eukaryot. Cell* **9,** 486–501 (2010).
64. Andrianov, V. *et al.* Tobacco as a production platform for biofuel: overexpression of Arabidopsis *DGAT* and *LEC2* genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnol. J.* **8,** 277–287 (2010).
65. La Russa, M. *et al.* Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. *J. Biotechnol.* http://dx.doi.org/10.1016/j.jbiotec.2012.04.006, (19 April 2012).
66. Keasling, J. D. & Chou, H. Metabolic engineering delivers next-generation biofuels. *Nature Biotechnol.* **26,** 298–299 (2008).
67. Lohr, M., Schwender, J. & Polle, J. E. Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Sci.* **185–186,** 9–22 (2012).
68. Eroglu, E. & Melis, A. Extracellular terpenoid hydrocarbon extraction and quantitation from the green microalgae *Botryococcus braunii* var. Showa. *Bioresour. Technol.* **101,** 2359–2366 (2010).
69. Wolf, F. R., Nemethy, E. K., Blanding, J. H. & Bassham, J. A. Biosynthesis of unusual acyclic isoprenoids in the alga *Botryococcus-braunii*. *Phytochemistry* **24,** 733–737 (1985).
70. Rosch, C., Skarka, J. & Wegerer, N. Materials flow modeling of nutrient recycling in biodiesel production from microalgae. *Bioresour. Technol.* **107,** 191–199 (2012).
71. Gilbert, N. Environment: the disappearing nutrient. *Nature* **461,** 716–718 (2009).
72. Yang, Y., Xu, J., Vail, D. & Weathers, P. *Ettlia oleabundans* growth and oil production on agricultural anaerobic waste effluents. *Bioresour. Technol.* **102,** 5076–5082 (2011).
73. Ras, M., Lardon, L., Bruno, S., Bernet, N. & Steyer, J. P. Experimental study on a coupled process of production and anaerobic digestion of *Chlorella vulgaris*. *Bioresour. Technol.* **102,** 200–206 (2011).
74. Zamalloa, C., Boon, N. & Verstraete, W. Anaerobic digestibility of *Scenedesmus obliquus* and *Phaeodactylum tricornutum* under mesophilic and thermophilic conditions. *Appl. Energy* **92,** 733–738 (2012).
75. Duan, P. G. & Savage, P. E. Upgrading of crude algal bio-oil in supercritical water. *Bioresour. Technol.* **102,** 1899–1906 (2011).
76. Levine, R. B., Pinnarat, T. & Savage, P. E. Biodiesel production from wet algal biomass through *in situ* lipid hydrolysis and supercritical transesterification. *Energy Fuels* **24,** 5235–5243 (2010).
77. Chen, Y. H. & Walker, T. H. Biomass and lipid production of heterotrophic microalgae *Chlorella protothecoides* by using biodiesel-derived crude glycerol. *Biotechnol. Lett.* **33,** 1973–1983 (2011).
78. Gachon, C. M. M., Sime-Ngando, T., Strittmatter, M., Chambouvet, A. & Kim, G. H. Algal diseases: spotlight on a black box. *Trends Plant Sci.* **15,** 633–640 (2010).
This report describes algal pathogens as highly diverse, and that although little is known about them, there are ecological benefits to their study.
79. Van Etten, J. L., Burbank, D. E., Kuczmarski, D. & Meints, R. H. Virus infection of culturable *Chlorella*-like algae and development of a plaque assay. *Science* **219,** 994–996 (1983).
80. Montagnes, D. J. S., Chambouvet, A., Guillou, L. & Fenton, A. Responsibility of microzooplankton and parasite pressure for the demise of toxic dinoflagellate blooms. *Aquat. Microb. Ecol.* **53,** 211–225 (2008).
81. Sukenik, A. *et al.* Photosynthetic performance of outdoor *Nannochloropsis* mass cultures under a wide range of environmental conditions. *Aquat. Microb. Ecol.* **56,** 297–308 (2009).
82. Natrah, F. M. I. *et al.* Effects of micro-algae commonly used in aquaculture on acyl-homoserine lactone quorum sensing. *Aquaculture* **317,** 53–57 (2011).
83. Murphy, C. F. & Allen, D. T. Energy-water nexus for mass cultivation of algae. *Environ. Sci. Technol.* **45,** 5861–5868 (2011).
84. Christenson, L. & Sims, R. Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts. *Biotechnol. Adv.* **29,** 686–702 (2011).
85. Wu, Z. *et al.* Evaluation of flocculation induced by pH increase for harvesting microalgae and reuse of flocculated medium. *Bioresour. Technol.* **110,** 496–502(2012).
86. Li, Q., Zhao, X. Q., Chang, A. K., Zhang, Q. M. & Bai, F. W. Ethanol-induced yeast flocculation directed by the promoter of *TPS1* encoding trehalose-6-phosphate synthase 1 for efficient ethanol production. *Metab. Eng.* **14,** 1–8 (2012).
87. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22,** 2971–2972 (2006).
88. Lan, E. I. & Liao, J. C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl Acad. Sci. USA* **109,** 6018–6023 (2012).
89. Melis, A., Zhang, L., Forestier, M., Ghirardi, M. L. & Seibert, M. Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*. *Plant Physiol.* **122,** 127–136 (2000).

Acknowledgements This work was supported by the US Department of Energy, grant number DE-EE-0003373, and the California Energy Commission Initiative for Large Molecule Sustainable Fuels, Agreement Number: 500-10-039.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of this paper at go.nature.com/1oa6yj. Readers are welcome to comment on the online version of this article at go.nature.com/1oa6yj. Correspondence should be addressed to S.P.M. (smayfield@ucsd.edu).

nature

www.nature.com/nature/supplements/insights

ARTICLE

doi:10.1038/nature11331

Passenger deletions generate therapeutic vulnerabilities in cancer

Florian L. Muller[1,2,3]*, Simona Colla[1,2,3]*, Elisa Aquilanti[2]*, Veronica E. Manzo[2], Giannicola Genovese[1,2], Jaclyn Lee[2], Daniel Eisenson[2], Rujuta Narurkar[2], Pingna Deng[1,2], Luigi Nezi[1,2], Michelle A. Lee[2,4], Baoli Hu[1,2,5], Jian Hu[1,2,3], Ergun Sahin[2,3], Derrick Ong[1,2,3], Eliot Fletcher-Sananikone[1,2], Dennis Ho[2,3], Lawrence Kwong[1,2], Cameron Brennan[6], Y. Alan Wang[1,2,5], Lynda Chin[1,2,2,5] & Ronald A. DePinho[2,3,5,7]

Inactivation of tumour-suppressor genes by homozygous deletion is a prototypic event in the cancer genome, yet such deletions often encompass neighbouring genes. We propose that homozygous deletions in such passenger genes can expose cancer-specific therapeutic vulnerabilities when the collaterally deleted gene is a member of a functionally redundant family of genes carrying out an essential function. The glycolytic gene enolase 1 (*ENO1*) in the 1p36 locus is deleted in glioblastoma (GBM), which is tolerated by the expression of *ENO2*. Here we show that short-hairpin-RNA-mediated silencing of *ENO2* selectively inhibits growth, survival and the tumorigenic potential of *ENO1*-deleted GBM cells, and that the enolase inhibitor phosphonoacetohydroxamate is selectively toxic to *ENO1*-deleted GBM cells relative to *ENO1*-intact GBM cells or normal astrocytes. The principle of collateral vulnerability should be applicable to other passenger-deleted genes encoding functionally redundant essential activities and provide an effective treatment strategy for cancers containing such genomic events.

Large-scale analysis of the cancer genome has provided an unprecedentedly detailed picture of the genetic anatomy of cancer[1], which has been, and continues to serve as, a blueprint for the development of molecular-targeted therapies. Targeted therapies directed against amplified or mutant-activated key driver oncoproteins have provided encouraging clinical progress[2], whereas exploiting loss-of-function mutations or gene deletions has received considerably less attention and has not been as successful thus far. Previous therapeutic work in the area of loss-of-function mutations and deletions has focused specifically on tumour-suppressor genes by strategies that include synthetic lethal approaches. One notable example of a synthetic lethal interaction is the response of BRCA1-mutant cancers to poly(ADP-ribose) polymerase inhibitors, although this interaction seems to be dependent on genetic context[3,4], and can be bypassed in late-stage tumours[3-5]. Most other synthetic lethal interactors of inactivated tumour suppressors seem to be less robust in eliciting cancer cell death[5], perhaps because such genes typically do not perform fundamental housekeeping functions.

Cancer genomes are characterized by numerous copy number amplifications and deletions, which target driver oncogenes and tumour-suppressor genes, respectively. Often, these genomic alterations are large regional events, affecting many other genes in addition to the intended target(s). The fact that such broad genomic alterations are not negatively selected against in cancer cells indicates that, on their own, the copy number alterations of these neighbouring passengers must not carry severely detrimental biological consequences. That said, it is conceivable that these passenger genomic events can create unintended (collateral) vulnerabilities unique to those cells; such as when a passenger being co-deleted is a member of a redundant multigene family serving an essential housekeeping function. A large body of genetic interaction studies in invertebrates as well as mice indicates that many essential cellular housekeeping functions are carried out by several homologous genes that encode overlapping functions; this redundancy enables cell viability after loss of one homologue but causes lethality after the loss of several homologues[6-10] (Supplementary Fig. 1). In this conceptual framework, we proposed that the homozygous deletion of redundant essential housekeeping genes could create cancer-specific vulnerabilities (Supplementary Fig. 1a), in which pharmacological inactivation of the second, non-deleted homologue would result in the complete loss of activity in tumour cells carrying the deletion, without compromising the health of normal cells, in which both genes are intact and expressed (Supplementary Fig. 1b).

ENO1 is a redundant housekeeping gene deleted in GBM

By examining The Cancer Genome Atlas (TCGA) GBM data set for homozygous deletions targeting genes involved in essential cell activities[1], we identified various such candidates, including the *ENO1* gene, which resides at the 1p36 tumour-suppressor locus (see Table 1 for a summary and Supplementary Table 1 for more detailed methodological support). Enolase, which is encoded by three homologous genes, is an essential enzyme that catalyses the second to last step of glycolysis, converting 2-phosphoglyceric acid into phosphoenolpyruvate[11]. In mammals, enolase activity is encoded by three genes: *ENO1*, which is ubiquitously expressed[12,13]; *ENO2*, which is expressed exclusively in neural tissues[12,14]; and *ENO3*, which is expressed in muscle tissues[15] (Supplementary Table 2). *ENO1* is the major enolase isoform in GBM, accounting for 75–90% of cellular enolase activity[12]. Given the crucial importance of glycolysis for energy generation and anabolic processes in normal and especially tumour cells[16], GBM tumours homozygous null for *ENO1* would be predicted to be highly sensitive to the inhibition of enolase 2, whereas

[1]Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3]Department of Genetics and Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [5]Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [6]Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [7]Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA.
*These authors contributed equally to this work.

Table 1 | Homozygously deleted essential-redundant genes in GBM

Homozygously deleted gene	Chromosomal locus	Proximal tumour-suppressor gene	Target homologue	Pathway
ENO1	1p36.2	Various	*ENO2*	Glycolysis and gluconeogenesis
H6PD	1p36.2	Various	*G6PD*	Pentose phosphate shunt
KIF1B	1p36.2	Various	*KIF1A/C*	Chromosomal segregation
NMNAT1	1p36.2	Various	*NMNAT2/3*	NAD^+ biosynthesis
UBE4B	1p36.2	Various	*UBE4A*	Polyubiquitin-dependent degradation
ACO1	9p21.1	*INK/ARF*	*ACO2/3*	Regulation of iron metabolism/citric acid cycle
KLHL9	9p22	*INK/ARF*	*KLHL13*	Chromosomal segregation
PANK1	10q23.31	*PTEN*	*PANK3*	Acetyl-CoA biosynthesis
KIF20B	10q23.31	*PTEN*	*KIF20A*	Chromosomal segregation/cytokinesis

An evidence-based filtered list of genes homozygously deleted in GBM that are likely to execute an essential housekeeping function and have redundant (and potentially druggable) homologues. A more detailed methods description is provided in Supplementary Table 1.

normal neural tissues should not be affected because of the functional redundancy of enolase 1 (Fig. 1a, b). Correspondingly, *Eno2* knockout mice are viable and fertile, suggesting that pharmacological inhibition of enolase 2 is likely to be well tolerated at the organism level (Supplementary Table 2). Moreover, *Saccharomyces cerevisiae*, which possesses several enolase homologues, shows weak phenotypes with single mutants and incurs cell lethality only when all homologues are deleted[9–15]; whereas *Caenorhabditis elegans* and *Drosophila* possess only one gene encoding enolase activity, and its deletion is lethal[17,18].

The 1p36 locus, which contains several candidate tumour-suppressor genes, including *CHD5* and *CAMTA1* (refs 19, 20), sustains frequent deletion in GBM (Fig. 2a). The 1p36 locus is homozygously deleted in 1–5% of GBMs[1,21,22] (as well as oligodendrogliomas[23] and large-cell neuroendocrine lung tumours[24]), and *ENO1* is often included in the deletion. By examining the TCGA copy number aberrations (single nucleotide polymorphism (SNP) and array comparative genomic hybridization (aCGH) data)[1] and expression profiles, we identified 5 out of 359 GBM samples with homozygous deletion of *ENO1* and associated near-complete absence of its expression (Fig. 2b and Supplementary Fig. 2). We identified two GBM cell lines, D423-MG[22] and Gli56 (ref. 25), with homozygous deletions at the 1p36 locus spanning *ENO1*. A third GBM cell line, D502-MG[22], also incurs homozygous deletion of many genes in this locus but not *ENO1*, thus serving as an excellent control (Fig. 2c). Western blot analysis confirmed the loss of ENO1 and the retention of ENO2 protein in D423-MG and Gli56 cells, whereas both proteins were present in D502-MG and in all other glioma and normal glial cell lines tested (Fig. 2d).

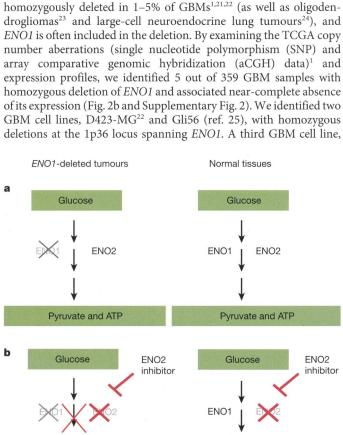


Figure 1 | Homozygous deletions in *ENO1* sensitize tumours to molecular targeting of ENO2. a, *ENO1* is homozygously deleted in glioblastomas as part of the 1p36 locus. Loss of ENO1 is tolerable to the tumour because ENO2 is still expressed. **b**, A specific ENO2 inhibitor should completely eliminate enolase activity in *ENO1*-null tumour cells (hence blocking glycolysis and ATP synthesis), but leave genomically intact normal tissues unaffected because enolase activity is still present as ENO1 is still expressed.

ENO2 knockdown inhibits growth of *ENO1*-deleted cells

We used the D502-MG (*ENO1* expressing) and D423-MG (*ENO1*-null) cell lines to assess the impact of short hairpin RNA (shRNA)-mediated knockdown of *ENO2* in an *ENO1* wild-type or null context. Two independent *ENO2* shRNAs (pLKO.1 vector) resulted in robust protein reduction and led to a profound inhibition of cell growth only in the context of *ENO1* genomic deletion (Supplementary Fig. 3). We obtained the same result using a further, independent *ENO2* shRNA (pGIPZ vector) (Supplementary Fig. 4a, b). Furthermore, shRNA ablation of *ENO2* in *ENO1*-null cells also resulted in decreased soft agar colony formation and blocked the *in vivo* tumorigenic potential of intracranially injected cells (Fig. 2e and Supplementary Fig. 4c). Finally, the selective toxicity of *ENO2* ablation to *ENO1*-null cells was demonstrated in an isogenic context using the doxycycline-inducible TRIPZ vector. When we used this doxycycline-inducible system in *ENO1* wild-type cell lines (U87, A1207 and LN319), two independent shRNAs reduced ENO2 protein levels by >70% (Fig. 3a) with no impact on ENO1 levels (data not shown). This *ENO2* ablation resulted in a profound inhibition of cell proliferation only in the *ENO1*-null D423-MG cell line (Fig. 3b). Furthermore, enforced expression of hairpin-resistant *ENO2* open reading frame fully reversed the deleterious effects of the *ENO2* shRNA (Supplementary Fig. 5), showing that the inhibitory effect of the hairpin was indeed specific to diminished ENO2 expression and was not an off-target effect. Finally, when ENO1 was ectopically re-expressed in D423-MG (*ENO1*-null) cell lines at levels similar to those observed in *ENO1* wild-type GBM lines, the deleterious effect of shRNA ablation of ENO2 was completely abrogated (Supplementary Fig. 6).

Enolase inhibition is toxic to *ENO1*-null cells

Next, we assessed the effect of pharmacological inhibition of enolase activity in *ENO1* wild-type and null cells. Previous studies have focused on the pharmacological inhibition of enolase, in particular for antiparasitic purposes[26,27], and many compounds have been characterized, most of which act as reaction-intermediate analogues (Supplementary Table 3). The most potent enolase inhibitor is phosphonoacetohydroxamate (PHAH)[27], which is thought to act as a transition-state analogue with an inhibitory constant of 15 pM on yeast enolase. Although PHAH has not been tested on human enolases, previous work demonstrated inhibitory effects on enolases from distantly related organisms[27,28], suggesting its potential use over a large phylogenetic distance. We find that PHAH was indeed capable of potent inhibition of enolase *in vitro* in native lysates of human GBM cell lines, with a half-maximum inhibitory concentration (IC_{50}) of around 20 nM (Fig. 4b and data not shown). We used PHAH in concentrations ranging from 0.625 μM to 50 μM and

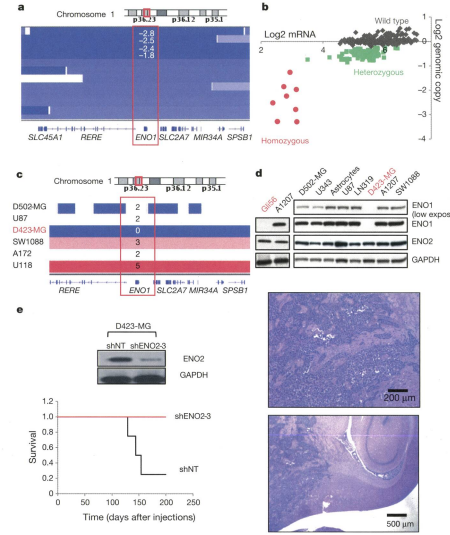


Figure 2 | Homozygous deletion of the 1p36 locus in GBM results in loss of *ENO1* expression in primary tumours and cell lines. a, TCGA aCGH data show four primary GBMs with log2 copy number <-1, indicating homozygous deletion of the 1p36 locus. **b**, DNA copy number correlates with messenger RNA expression; expression is highest in tumours with $n = 2$ copies (wild type) and lowest in tumours with $n = 0$ copies (null) of *ENO1*. **c**, The D423-MG cell line was identified as homozygously deleted by SNP arrays from the Wellcome Trust Sanger Institute data set. **d**, The complete absence of ENO1 protein in D423-MG and Gli56 cells was confirmed by western blotting. **e**, shRNA targeting *ENO2* (shENO2-3) but not scramble (shNT) in D423-MG *ENO1*-null cells ablated intracranial tumorigenesis *in vivo* ($n = 4$ mice per group).

observed marked toxicity in *ENO1*-null cells (Fig. 4a, c and Supplementary Fig. 7) and minimal impact on the *ENO1* wild-type controls, which show at least ten times greater enolase activity relative to the *ENO1*-null cells (because ENO1 accounts for 90% of the total cellular enolase activity[12]; Fig. 4b). Although the IC_{50} of PHAH is similar for ENO1 and ENO2 *in vitro* (data not shown) the greater toxicity of the inhibitor to *ENO1*-null cells (Gli56 and D423-MG) derives from the fact that in these cells enolase activity is already 90% lower than in wild-type ENO1 cell lines, and consequently, a much lower dose is required to decrease total enolase activity below the toxicity threshold. Further data indicate a direct relationship between the levels of enolase activity and the sensitivity to PHAH across different cell lines and in the same cell line with different levels of enforced enolase expression. First, U343 and D502-MG cells, which have intermediate levels of enolase activity (and ENO1 protein expression; Fig. 2d) compared with the other cell lines, have intermediate levels of sensitivity to PHAH (Fig. 4), which in the case of U343 can be rescued by ectopic overexpression of *ENO1* or *ENO2* (data not shown). A systematic titration of PHAH in D423-MG cell lines with varying levels of enforced ENO1 or ENO2 expression, shows a direct relationship between the level of enolase expression/activity and the ensuing resistance to PHAH (Supplementary Fig. 7). PHAH toxicity was also abrogated in Gli56 *ENO1*-null cells by ectopic expression of physiological levels of *ENO1* or overexpression of *ENO2* (Supplementary Fig. 7). Regarding the mechanism of toxicity, cell cycle and apoptosis analysis demonstrated that PHAH treatment for 48 h induced a marked decrease of S-phase cells followed by a marked increase of apoptosis in D423-MG but not in *ENO1* wild-type U373 cells (Supplementary Table 4). This effect was completely rescued by *ENO2* overexpression (data not shown). The fact that this growth inhibition and subsequent apoptosis is due to energy crisis is substantiated by a strong induction of phosphorylated 5′-AMP-activated protein kinase (AMPK) (at Thr 172)[29], which was observed in D423-MG but not in ENO1 wild-type cell lines (data not shown). It is tempting to speculate that this energy stress response exerts a protective effect, and thus the concomitant addition of an AMPK inhibitor together with PHAH could result in further toxicity. Finally, it is worth noting that *ENO1*-null cells do not show any greater sensitivity to other molecular-targeted therapies, such as a combination of receptor tyrosine kinase inhibitors[30] (lapatinib, sorafenib and PHA665752) (Supplementary Fig. 8) and rapamycin (data not shown) compared with wild-type *ENO1* cells. These data indicate that D423-MG cells are not broadly susceptible to other anticancer agents and that PHAH selectively targets *ENO1*-null GBM cells.

Discussion

In this study, we sought to determine the effect of collateral deletion of genes in tumour-suppressor loci that belong to redundant gene families playing cell-essential roles, and to assess whether extinction of remaining gene family members would create cancer-specific vulnerabilities. We provide genetic and pharmacological evidence that enolase 2 inhibition is lethal in cells with 1p36 homozygous deletion with collateral loss of *ENO1*, whereas *ENO1*-intact cells can rely on

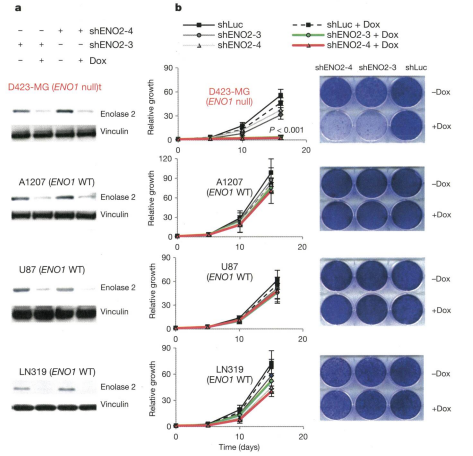


Figure 3 | shRNA ablation of *ENO2* affects *ENO1*-null but not *ENO1*-WT GBM cells. **a**, shRNA ablation by two independent doxycycline (Dox)-inducible TRIPZ hairpins against *ENO2* (shENO2-3, shENO2-4) resulted in >70% reduction in enolase 2 protein levels in both *ENO1* wild-type (WT) (A1207, U87 and LN319) and *ENO1*-null (D423-MG) cell lines. **b**, Ablation of *ENO2* markedly inhibited growth of *ENO1*-null but not wild-type *ENO1* cells, whereas non-targeting shRNA against luciferase (shLuc) had no effect in any cell line ($n = 3$ biological replicates, mean ± s.e.m., Student's *t*-test). Representative plates at the last time point of growth for cells infected with shLuc, shENO2-3 or shENO2-4, with or without Dox induction, are shown to the right of growth curves for each cell line.

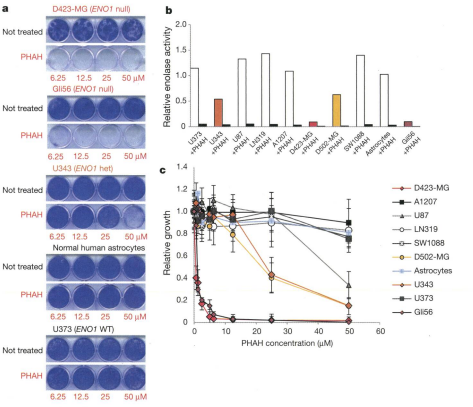


Figure 4 | Extreme sensitivity of *ENO1*-null cells to the pan-enolase inhibitor PHAH. **a**, D423-MG and Gli56 *ENO1*-null cell lines are highly sensitive to PHAH toxicity, whereas *ENO1* wild-type cell lines and normal astrocytes are not. Het, heterozygous deleted. **b**, The sensitivity of GBM lines to PHAH treatment correlated with their overall enolase activity. Pre-incubation of the lysates with 1 μM PHAH inhibited enolase enzymatic activity by >95% (average, $n = 2$ technical replicates). **c**, PHAH minimally affected the growth of *ENO1* wild-type GBM cells and normal astrocytes except at concentrations above 50 μM. Low PHAH concentrations stall the growth of *ENO1*-null cells, whereas cells with reduced *ENO1* expression (D502-MG and U343) showed intermediate sensitivity ($n = 4$ biological replicates, mean ± s.e.m.).

enolase 1 to undergo glycolysis and support survival. These findings are in agreement with genetic data from invertebrates[8–10]. Given that several homozygously deleted housekeeping genes can occur in the same deletion on 1p36 (for example, *H6PD*; Supplementary Table 1), it may be possible to increase the effectiveness and cancer cell-specific killing further by combining the inhibition of *ENO2* with that of another homologue of a simultaneously deleted housekeeping gene.

Attempts to therapeutically exploit general metabolic differences between normal and cancer cells, such as glucose addiction (the Warburg effect) and glutamine or serine addiction[16,31,32], remain areas of active preclinical investigation and clinical development. The approach described here is distinguished from these attempts in that it does not rely on any general aspect of cancer cell metabolism, but rather rests on genetically determined metabolic differences between normal and cancerous tissue to generate cancer cell-specific vulnerabilities. We propose that collateral vulnerability may be extended to other passenger homozygously deleted housekeeping genes in loci sustaining frequent deletion, such as 9p21 (*CDKN2A*) and 10q23 (*PTEN*), which contain members of functionally redundant housekeeping gene families (Table 1, Supplementary Table 1 and Supplementary Fig. 9). The strongest pharmacologically targetable candidates in this list are the homologues of the dual-function energy metabolism/iron regulator aconitase 1 (*ACO1*; 9p21) and the coenzyme A biosynthetic enzyme, pantothenate kinase 1 (*PANK1*; 10q23). Importantly, many of the compounds targeting the homologues of these passenger genes represent new molecular entities with respect to cancer treatment.

By one estimate, 11% of all protein-coding genes in the human genome are deleted in human cancers[33]. Thus, given the large number of homozygous deletions across many different cancer types spanning many hundreds of genes[33–36], the model described here for GBM should be applicable to the development of personalized treatments for many other cancer types.

METHODS SUMMARY

Cells were cultured using standard techniques in DMEM with 20% fetal bovine serum. shRNA experiments were conducted by lentiviral production through transient transfection of 293T cells followed by transduction in medium containing 4 $\mu g\,ml^{-1}$ polybrene and selection with 2 $\mu g\,ml^{-1}$ puromycin. shRNA expression was induced with 1 $\mu g\,ml^{-1}$ doxycycline and knockdown was tested by western blot. The shRNA-resistant ENO2 open reading frame was created by introducing silent mutations with the QuickChange site directed mutagenesis kit from Stratagene and then cloning into the pHAGE-CMV lentiviral vector. Cell proliferation experiments were performed using crystal violet staining, the CellTiter-Glo assay (Roche) and by measuring confluence using IncuCyte (Essen Bioscience). Orthotopic intracranial injections of D423-MG cells with and without ENO2 knockdown in severe combined immunodeficient (SCID) mice were performed as previously described[37]. Soft agar colony formation assay of these cells was performed using standard techniques by seeding 10^4 cells in 6-well plates. For the inhibitor studies, PHAH lithium salt was custom-synthesized by TCRS, following a previous protocol[38]. For the enolase activity assay, NADH oxidation was measured in a pyruvate kinase–lactate dehydrogenase coupled reaction as previously described[21]. For cell cycle studies, cells were incubated with or without PHAH for 48 h, stained with propidium iodide and sorted by flow cytometric analysis. For annexin V/7-AAD assays, cells were treated with or without PHAH for 96 h, stained with annexin V–phycoerythrin (PE) and 7-AAD, and evaluated for apoptosis by flow cytometry according to the manufacture's protocol (Biovision).

Full Methods and any associated references are available in the online version of the paper.

Received 16 January; accepted 15 June 2012.

1. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
2. Druker, B. J. Translation of the Philadelphia chromosome into therapy for CML. *Blood* **112,** 4808–4817 (2008).
3. Guha, M. PARP inhibitors stumble in breast cancer. *Nature Biotechnol.* **29,** 373–374 (2011).
4. De Soto, J. A. & Deng, C. X. PARP-1 inhibitors: are they the long-sought genetically specific drugs for BRCA1/2-associated breast cancers? *Int. J. Med. Sci.* **3,** 117–123 (2006).
5. Weidle, U. H., Maisel, D. & Eick, D. Synthetic lethality-based targets for discovery of new cancer therapeutics. *Cancer Genomics Proteomics* **8,** 159–171 (2011).
6. Vavouri, T., Semple, J. I. & Lehner, B. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet.* **24,** 485–488 (2008).
7. Brookfield, J. F. Genetic redundancy. *Adv. Genet.* **36,** 137–155 (1997).
8. Castanzo, M. *et al.* The genetic landscape of a cell. *Science* **327,** 425–431 (2010).
9. DeLuna, A. *et al.* Exposing the fitness contribution of duplicated genes. *Nature Genet.* **40,** 676–681 (2008).
10. Deutscher, D., Meilijson, I., Kupiec, M. & Ruppin, E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature Genet.* **38,** 993–998 (2006).
11. Poyner, R. R. & Reed, G. H. Structure of the bis divalent cation complex with phosphonoacetohydroxamate at the active site of enolase. *Biochemistry* **31,** 7166–7173 (1992).
12. Joseph, J., Cruz-Sanchez, F. F. & Carreras, J. Enolase activity and isoenzyme distribution in human brain regions and tumors. *J. Neurochem.* **66,** 2484–2490 (1996).
13. Stefanini, M. Chronic hemolytic anemia associated with erythrocyte enolase deficiency exacerbated by ingestion of nitrofurantoin. *Am. J. Clin. Pathol.* **58,** 408–414 (1972).
14. Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450,** 503–508 (2007).
15. Comi, G. P. *et al.* β-enolase deficiency, a new metabolic myopathy of distal glycolysis. *Ann. Neurol.* **50,** 202–207 (2001).
16. Wise, D. R. & Thompson, C. B. Glutamine addiction: a new therapeutic target in cancer. *Trends Biochem. Sci.* **35,** 427–433 (2010).
17. Buszczak, M. *et al.* The carnegie protein trap library: a versatile tool for *Drosophila* developmental studies. *Genetics* **175,** 1505–1531 (2007).
18. Sonnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434,** 462–469 (2005).
19. Henrich, K. O. *et al.* CAMTA1, a 1p36 tumor suppressor candidate, inhibits growth and activates differentiation programs in neuroblastoma cells. *Cancer Res.* **71,** 3142–3151 (2011).
20. Bagchi, A. & Mills, A. A. The quest for the 1p36 tumor suppressor. *Cancer Res.* **68,** 2551–2556 (2008).
21. Yin, D. *et al.* High-resolution genomic copy number profiling of glioblastoma multiforme by single nucleotide polymorphism DNA microarray. *Mol. Cancer Res.* **7,** 665–677 (2009).
22. Duncan, C. G. *et al.* Integrated genomic analyses identify ERRFI1 and TACC3 as glioblastoma-targeted genes. *Oncotarget* **1,** 265–277 (2010).
23. Kotliarov, Y. *et al.* High-resolution global genomic survey of 178 gliomas reveals novel regions of copy number alteration and allelic imbalances. *Cancer Res.* **66,** 9428–9436 (2006).
24. Peng, W. X. *et al.* Array-based comparative genomic hybridization analysis of high-grade neuroendocrine tumors of the lung. *Cancer Sci.* **96,** 661–667 (2007).
25. Mueller, W. *et al.* Downregulation of RUNX3 and TES by hypermethylation in glioblastoma. *Oncogene* **26,** 583–593 (2005).
26. Guha-Chowdhury, N., Clark, A. G. & Sissons, C. H. Inhibition of purified enolases from oral bacteria by fluoride. *Oral Microbiol. Immunol.* **12,** 91–97 (1997).
27. de A. S. & Navarro, M. V. *et al.* Structural flexibility in *Trypanosoma brucei* enolase revealed by X-ray crystallography and molecular dynamics. *FEBS J.* **274,** 5077–5089 (2007).
28. Anderson, V. E., Weiss, P. M. & Cleland, W. W. Reaction intermediate analogues for enolase. *Biochemistry* **23,** 2779–2786 (1984).
29. Jing, M. & Ismail-Beigi, F. Critical role of 5′-AMP-activated protein kinase in the stimulation of glucose transport in response to inhibition of oxidative phosphorylation. *Am. J. Physiol. Cell Physiol.* **292,** C477–C487 (2007).
30. Stommel, J. M. *et al.* Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318,** 287–290 (2007).
31. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476,** 346–350 (2011).
32. Raj, L. *et al.* Selective killing of cancer cells by a small molecule targeting the stress response to ROS. *Nature* **475,** 231–234 (2011).
33. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463,** 893–898 (2010).
34. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18,** 11–22 (2010).
35. Cox, C. *et al.* A survey of homozygous deletions in human cancer genomes. *Proc. Natl Acad. Sci. USA* **102,** 4542–4547 (2005).
36. Tonon, G. *et al.* High-resolution genomic profiles of human lung cancer. *Proc. Natl Acad. Sci. USA* **102,** 9625–9630 (2005).
37. Zheng, H. *et al.* p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation. *Nature* **455,** 1129–1133 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Ligon, C. Maire, D. N. Louis, J. Kim and G. Mohapatra for sharing bioinformatics data from their tumour neurosphere banks. We also thank D. Bigner for sharing the D423-MG and D502-MG cell lines and D. N. Louis and J. Kim for sharing the Gli56 cell line. We thank G. Chu and D. Jakubosky for assistance with necropsy and histopathological analysis. F.L.M. was supported by a training grant from

the National Institutes of Health (NIH T32-CA009361) and a fellowship from the American Cancer Society (115992-PF-08-261-01-TBE). S.C. was supported by a Dana-Farber Cancer Institute/Harvard Cancer Center Myeloma SPORE career development grant. E.A. was supported by a Howard Hughes Medical Institute Medical Research Fellowship (57006984). V.M. was supported by a Harvard PRISE fellowship. M.A.L. was supported by a Diversity in Health-Related research award (3 P01 CA095616-0851). F.L.M. thanks J. Mohr for assistance with figure preparations. We also thank K. Muller for assistance with manuscript editing. We thank all members of the DePinho and Chin laboratories for suggestions and discussions. This work is supported by the NIH (P01CA95616 to C.B., L.C. and R.A.D.) and by the Ben and Catherine Ivy Foundation (to R.A.D. and L.C.).

Author Contributions F.L.M. and R.A.D. generated the original hypothesis. F.L.M. performed all bioinformatics work, including scanning the TCGA data set (with initial assistance from J.H.) and identifying candidates for collateral lethality, with the exception of *KLHL9*, which was identified by E.F.-S. E.A. obtained the D423-MG cell line and designed and carried out the pLKO and pGIPZ shRNA experiments. S.C. designed and performed all shRNA experiments with the inducible vectors and rescue experiments. F.L.M. and E.A. identified PHAH, F.L.M. procured the compound, and F.L.M. and S.C. performed all inhibitor treatment experiments. L.N. generated shRNA-resistant constructs of *ENO2*. S.C., D.O. and E.F.-S. performed cell cycle and apoptosis assays. R.N., V.M., D.E., P.D. and J.L. performed cell culture, crystal violet staining, western blotting and associated experiments and assisted in the preparation of figures. C.B. provided extensive unpublished genomic data and reagents from his primary brain tumour and neurosphere bank for Supplementary Table 1. E.A., M.A.L., B.H. and G.G. performed tumour cell injections. D.H., E.S., L.K., Y.A.W. and L.C. provided intellectual contributions throughout the project. F.L.M., E.A., S.C., Y.A.W., L.C. and R.A.D. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.D. (rdepinho@mdanderson.org).

ARTICLE

doi:10.1038/nature11347

Division and subtraction by distinct cortical inhibitory networks *in vivo*

Nathan R. Wilson[1]*, Caroline A. Runyan[1]*, Forea L. Wang[1] & Mriganka Sur[1]

Brain circuits process information through specialized neuronal subclasses interacting within a network. Revealing their interplay requires activating specific cells while monitoring others in a functioning circuit. Here we use a new platform for two-way light-based circuit interrogation in visual cortex *in vivo* to show the computational implications of modulating different subclasses of inhibitory neurons during sensory processing. We find that soma-targeting, parvalbumin-expressing (PV) neurons principally divide responses but preserve stimulus selectivity, whereas dendrite-targeting, somatostatin-expressing (SOM) neurons principally subtract from excitatory responses and sharpen selectivity. Visualized *in vivo* cell-attached recordings show that division by PV neurons alters response gain, whereas subtraction by SOM neurons shifts response levels. Finally, stimulating identified neurons while scanning many target cells reveals that single PV and SOM neurons functionally impact only specific subsets of neurons in their projection fields. These findings provide direct evidence that inhibitory neuronal subclasses have distinct and complementary roles in cortical computations.

Inhibition has fundamental and diverse roles in brain function, and is delivered by specialized cell types with distinct intrinsic properties and connectivity patterns[1–3]. This heterogeneity in cellular form and function suggests that different inhibitory subtypes may actually underpin distinct computational functions and even hold specific relevance to neurological disorders[4] based on their unique morphologies and functional positions within the network. Previous pharmacological or intracellular studies in primary visual cortex (V1), which necessarily considered inhibition as a single entity, have produced diverse findings on the role of inhibition. On the one hand, inhibition has been proposed to sharpen neuronal responses by removing weak inputs[5–7], though there have been conflicting reports on whether inhibition predominantly targets non-preferred responses or preferred ones[8–10]. On the other hand, inhibition has been posited to control response gain, a network mechanism by which cortical networks rapidly 'divide' or scale their dynamic range of responses[11]. This mechanism has been proposed as fundamental to processing across many brain systems, from primary sensory computations[12] to attention[13], multisensory integration[14], and value estimation[15].

Here we show that inhibition in the cerebral cortex can have either of these functions, depending on its cellular source. We propose that two key inhibitory neuron subclasses, soma-targeting PV neurons and dendrite-targeting SOM neurons, which together comprise a substantial proportion of cortical inhibitory neurons in mice[16,17], drive different kinds of inhibition. We combined optogenetic activation of individual or populations of PV or SOM neurons with monitoring the effects in target cells using high-speed imaging of functional responses[18] as well as cell-attached electrophysiological recordings[19]. These methods complement both static wiring diagrams[20] and wiring patterns examined in tissue slices[21–24] by revealing targeting specificity and functional consequences of inhibitory neuron activation in intact circuits processing visual information.

Optical dissection of network interactions

To measure the effects of distinct cell classes within a functioning network, we built a custom system combining optogenetic stimulation with *in vivo* two-photon imaging in the mammalian brain (Fig. 1a and Supplementary Fig. 1). Our imaging system (Supplementary Movie 1) sampled calcium responses from neurons loaded with a fluorescent reporter using a scan path customized for each image[25], at high speed but also high dwell times within neurons, yielding highly repeatable measurements of orientation-selective responses and clear tuning curves (Fig. 1a–f, Supplementary Figs 2 and 3, and Supplementary Movie 2).

To optically activate PV or SOM neurons, in parallel experiments, we used Cre/*loxP* recombination to express channelrhodopsin-2 (ChR2) in PV or SOM neurons in the mouse visual cortex (Supplementary Fig. 4). This led to highly specific and reliable on-demand activation of infected neurons in visual cortex that was verified both in slices and during visual stimulation *in vivo* (Fig. 1g and Supplementary Fig. 4). PV or SOM neurons were photo-activated for a 1-s interval at the onset of visual stimulation (Fig. 1h), enabling us to compare the control visual responses of neighbouring, non-infected neurons during episodically presented drifting oriented gratings (Fig. 1i) to responses in interleaved trials in which the cells were inhibited through PV or SOM activation (Fig. 1j and Supplementary Fig. 5). Concurrent calcium imaging and optogenetic stimulation enabled us to quantify interneuron suppression of neighbouring cells across the network (Fig. 1k).

Distinct functions of inhibitory cell classes

Using this system, we activated PV or SOM cells while recording the visual responses of non-infected cells to oriented drifting gratings (Supplementary Movie 3, Fig. 2a–d and Supplementary Fig. 5). Control responses of target cells were similar in PV and SOM experiments (Supplementary Fig. 6), and suppression by ChR2 was calibrated to a similar moderate range for all experiments (Supplementary Fig. 6i). Interestingly, PV activation caused a larger suppression when control responses were higher (Fig. 2e), whereas SOM activation caused a relatively uniform suppression of the full response profile (Fig. 2f), particularly when the control responses or baseline levels were high relative to the suppression thereby avoiding a 'floor effect'. Indeed,

[1]Department of Brain and Cognitive Sciences, Picower Institute for Learning and Memory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, 02139 USA.
*These authors contributed equally to this work.

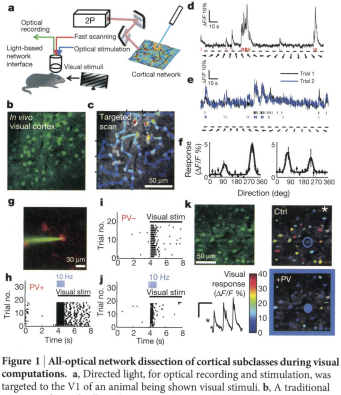


Figure 1 | All-optical network dissection of cortical subclasses during visual computations. **a**, Directed light, for optical recording and stimulation, was targeted to the V1 of an animal being shown visual stimuli. **b**, A traditional raster scan located cells in the network that were bulk-loaded with the calcium indicator dye OGB. **c**, Cells that were automatically identified were then imaged at high speed using targeted two-photon scanning along an arbitrary scan path. **d**, This enabled detection of robust cellular activity in response to episodically presented oriented drifting gratings, and this activity could be analysed as primary signals or deconvolved to estimate action potentials (red lines). **e**, Evoked optical traces were highly consistent over repeated presentations of visual stimuli. **f**, The clear, repeatable responses enabled the resolution of well-defined responses (dots), fit by dual Gaussian curves (lines). Data are shown as mean ± s.e.m. **g**, Image of an mCherry-ChR2+ PV+ (channelrhodopsin-2-and PV-positive) cell targeted *in vivo* for cell-attached recording. **h**, Evoked action potentials from an mCherry-ChR2+ PV+ cell. PV+ neurons were activated at the onset (3.9 s) of 4 s of visual stimulation, through 10-Hz stimulation of the PV+ neuron. **i**, **j**, Recording of the visual response of a PV− neuron in the control condition (**i**) and with PV activation (**j**). **k**, Left, cell population loaded with OGB dye (top), and responses of a cell marked by an asterisk in the network shown in the top right panel (bottom). Right, cells colour-coded by the magnitude of their visual response in the control condition (top) and when PV+ neurons were activated (bottom; colour bar, $\Delta F/F$ %). A ChR2+ PV neuron in the network is circled. +PV, optical PV activation; Ctrl, control; PV−, PV-negative; PV+, PV-positive.

the slope of the relationship between relative suppression and control response was significantly greater when PV cells were activated (Fig. 2g, i) than when SOM cells were activated (Fig. 2h, i; PV, $n = 150$ cells; SOM, $n = 77$ cells; $P < 0.001$, Kolmogorov–Smirnov test). Comparing the suppression at different response strengths for pooled cells further showed the asymmetric relationship between response and suppression when PV cells were activated (Fig. 2j; responses at 40, 60 and 80% of the maximum were less suppressed than those at maximum response strength; $P < 0.05$–0.001 for all pair-wise comparisons with 100%). In contrast, SOM suppression affected weak and strong responses similarly (Fig. 2k; responses at 40, 60 and 80% compared to responses at 100%; $P > 0.2$). Thus, suppression by SOM cells is relatively uniform across responses of different strengths, whereas suppression by PV cells is non-uniform and proportional to the response level of the target neuron.

Proportionate suppression would 'scale' responses, reducing the tuning curve's peak more strongly than its spontaneous response level or baseline, whereas uniform suppression would 'shift' the entire tuning curve downwards, including the baseline (Supplementary Fig. 7). Indeed, averaging the tuning curves of cells recorded during PV-cell ChR2 activation revealed a scaled down version of the control tuning curve (Fig. 2l), whereas activating SOM-cell ChR2 yielded a more uniform downwards shift of the control curve (Fig. 2m).

The presence of a response 'floor' or threshold influences how different forms of suppression impact the orientation tuning curve (Supplementary Fig. 7). Examining cells in which post-ChR2 responses were largely above the 'floor', to accurately quantify the full distribution of suppression across the whole curve (Supplementary Fig. 7a), showed that PV and SOM activation both reduced the baseline responses at non-preferred orientations (Fig. 2n; +PV (PV neurons activated optically), 11 ± 4% decrease in baseline, $P < 0.05$, $n = 52$ cells; versus +SOM, 19 ± 6%, $P < 0.001$, $n = 25$ cells). PV but not SOM activation decreased the peak-baseline amplitude of cells' tuning functions (Fig. 2o; +PV, 28.5 ± 4.5% reduction, $P < 0.001$; +SOM, 19.8 ± 11.2%, $P = 0.10$), consistent with a larger reduction at the peak relative to the baseline. PV activation did not affect the orientation selectivity index (OSI) of target cells (Fig. 2p; control, 0.29 ± 0.01; versus +PV, 0.29 ± 0.01; $P = 0.71$; see also Supplementary Fig. 8), whereas SOM activation increased the OSI (Fig. 2p; control, 0.26 ± 0.01; versus +SOM, 0.30 ± 0.01; $P < 0.01$). Similarly, PV activation did not affect the tuning width of target cells (Fig. 2q; control, half-width at half-height, 38.7 ± 3.0 degrees; versus +PV, 35.2 ± 3.0 deg; $P = 0.31$), whereas SOM activation narrowed the tuning width (Fig. 2q; control, 45.7 ± 4.1 deg; versus +SOM, 37.5 ± 4.3 deg; $P < 0.01$). The effects on the direction selectivity index (DSI) were similar in trend (Fig. 2r; PV: 0.33 ± 0.03 control DSI versus 0.36 ± 0.03 +PV DSI, $P = 0.44$; SOM: 0.28 ± 0.04 control DSI versus 0.34 ± 0.05 +SOM DSI, $P = 0.10$). Thus, the relatively uniform suppression by SOM cells leads to a sharpening in response selectivity of target neurons, whereas the non-uniform but proportional suppression by PV cells reduces response magnitude but does not change response selectivity.

PV and SOM effects measured electrophysiologically

We examined further the different effects of PV and SOM activation using electrophysiological cell-attached recordings *in vivo* (Fig. 3a). Putative pyramidal neurons (Fig. 3b) were identified by their regular spiking properties (Fig. 3c); the peak:trough ratio of individual spikes was larger for all recorded cells than in identified fast-spiking PV-positive neurons (2.82 ± 0.20 for all recorded cells, $n = 21$; 1.41 ± 0.18 in PV-positive neurons, $n = 53$; $P < 0.001$). Moderate levels of PV- and SOM-mediated suppression (Supplementary Fig. 6i) had clearly different effects on spike responses of target cells and resultant orientation tuning curves (Fig. 3d, e), even in cells with very different response levels (Fig. 3f, g). PV suppression depended on the level of control response, whereas SOM suppression shifted tuning curves downwards more uniformly (Fig. 3h–k; PV, $n = 21$ cells, SOM, $n = 17$ cells). The changes in average tuning curves (Fig. 3l, m), and their parameters (Fig. 3n–r; Supplementary Fig. 8), showed that SOM but not PV suppression sharpened response selectivity, consistent with a model in which PV activation leads to a division of target cell responses but in which SOM activation leads to a subtraction (Supplementary Fig. 9).

Differential inhibitory impact on target–cell gain

The divisive impact of PV activation suggests that PV neurons implement dynamic response gain control in cortex, which has been previously attributed to intracortical inhibition[26,27]. We carried out cell-attached recordings and examined the effects of PV and SOM activation on a canonical measure of response gain, the modulation of responses with increasing contrast (Fig. 4a, b). PV activation (Fig. 4c) led to contrast response curves with reduced gain (slope), whereas SOM activation (Fig. 4d) decreased responses relatively uniformly (with a floor effect at low response levels). The PV activation curves

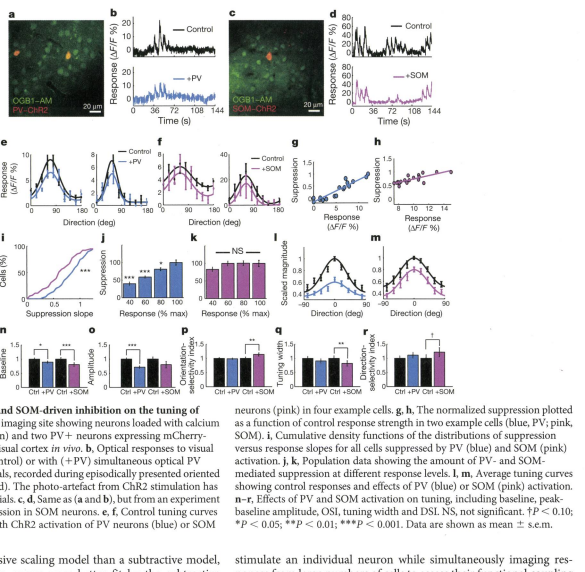


Figure 2 | Impact of PV- and SOM-driven inhibition on the tuning of neuronal responses. **a**, An imaging site showing neurons loaded with calcium indicator (OGB1-AM, green) and two PV+ neurons expressing mCherry-ChR2 (PV-ChR2, red) in visual cortex *in vivo*. **b**, Optical responses to visual stimuli, either without (Control) or with (+PV) simultaneous optical PV activation in interleaved trials, recorded during episodically presented oriented drifting gratings (see Fig. 1d). The photo-artefact from ChR2 stimulation has been removed from these trials. **c**, **d**, Same as (**a** and **b**), but from an experiment with mCherry-ChR2 expression in SOM neurons. **e**, **f**, Control tuning curves (black) were suppressed with ChR2 activation of PV neurons (blue) or SOM neurons (pink) in four example cells. **g**, **h**, The normalized suppression plotted as a function of control response strength in two example cells (blue, PV; pink, SOM). **i**, Cumulative density functions of the distributions of suppression versus response slopes for all cells suppressed by PV (blue) and SOM (pink) activation. **j**, **k**, Population data showing the amount of PV- and SOM-mediated suppression at different response levels. **l**, **m**, Average tuning curves showing control responses and effects of PV (blue) or SOM (pink) activation. **n**–**r**, Effects of PV and SOM activation on tuning, including baseline, peak-baseline amplitude, OSI, tuning width and DSI. NS, not significant. †$P < 0.10$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Data are shown as mean ± s.e.m.

were better fit by a divisive scaling model than a subtractive model, whereas SOM activation curves were better fit by the subtractive model (Supplementary Fig. 10). PV- but not SOM-mediated suppression was dependent on the response level, as shown by both suppression-response slopes (Fig. 4e, f; PV, 0.11 ± 0.03, $n = 17$ cells; versus SOM, −0.008 ± 0.02, $n = 16$ cells; $P < 0.01$; Supplementary Fig. 11a, b) and suppression-response strength comparisons (PV, $P < 0.05$–0.001 comparing suppression at 40%, 60% and 80% response to that at 100% (Fig. 4g); SOM, $P > 0.2$ comparing suppression at 40%, 60% and 80% response to that at 100% (Fig. 4h)). PV activation scaled response magnitude (Fig. 4i; +PV: R_{max} 64.1 ± 3.5% of control; $n = 17$ cells; $P < 0.001$) without affecting half-saturation contrast (Fig. 4j; +PV: C_{50} 95.0 ± 10% of control; $P = 0.65$). SOM activation significantly reduced R_{max} (Fig. 4i; +SOM: 73.6 ± 5.3% of control; $n = 16$ cells; $P < 0.001$), but unlike PV, also significantly increased C_{50} (Fig. 4j; +SOM: 141 ± 17.7% of control; $P < 0.05$), with no effect on the response gain ($P = 0.23$ comparing slopes at C_{50} before and after SOM activation; versus $P < 0.01$ comparing slopes before and after PV activation; see also Supplementary Fig. 11). Thus, PV but not SOM activation contributes directly to controlling the gain of target-cell responses.

Single-cell circuit maps of network connections *in vivo*

The functional roles of inhibitory neurons are manifested through the spatial distribution and functional targeting of subclasses of inhibition onto cortical cells in the local network. To define the output connections of single inhibitory neurons, we developed a system to focally stimulate an individual neuron while simultaneously imaging responses from large numbers of cells to assess their functional coupling (Fig. 5a). The ChR2-stimulating 473-nm beam was narrowed to a small effective radius and focused on sparsely distributed ChR2-positive cells (Fig. 5b and Supplementary Figs 12–16). Thus, we could optically activate chosen PV or SOM neurons *in vivo* while concurrently sampling population responses with targeted imaging (Fig. 5c and Supplementary Movie 4).

Mapping response modulation across a network while controlling a PV ChR2 cell or a SOM ChR2 cell (Fig. 5d, h), we found that visual responses of some neighbouring neurons were significantly suppressed while other cells were unaffected (Fig. 5e, i), resulting in maps of the functional suppression triggered during focal PV or SOM activation (Fig. 5f, j). Similar non-uniform, heterogeneous maps of functional PV and SOM cell connectivity were obtained from every imaged animal (Supplementary Fig. 17; PV, $n = 4$ networks; SOM, $n = 5$ networks). Focal PV activation resulted in the significant suppression of 43.1 ± 2.1% of neurons within the field of view, whereas focal SOM activation suppressed 16.2 ± 2.9% of neurons. Electrical stimulation of a single cell through whole-cell patch recording *in vivo* yielded a similar suppression map (Supplementary Fig. 18). Single PV or SOM neuron activation rarely triggered observable dis-inhibition (Supplementary Fig. 19). The nature of suppression was very similar to that observed with full-field activation: the amount of suppression by PV cells depended on the strength of the control response, whereas focal SOM stimulation resulted in more uniform suppression (Fig. 5h, m; see also Supplementary Fig. 20). Thus, the effects of SOM- and

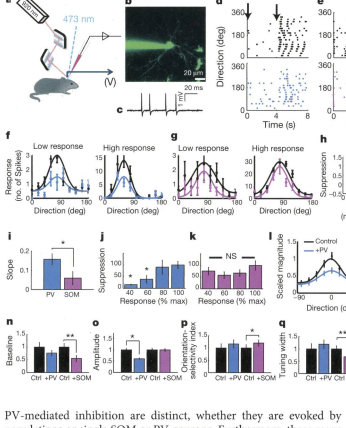


Figure 3 | Electrophysiological analysis of PV- and SOM-driven inhibition. **a**, PV or SOM neurons were optogenetically activated while recording visual responses (measured in volts (V)) using targeted *in vivo* cell-attached recordings. **b**, A pyramidal neuron was patched under two-photon guidance and filled with Alexa 488 dye. **c**, Representative spikes recorded. **d**, **e**, Spike rasters depict responses of target cells in time over 18 directions, pooled over multiple trials in control conditions (black) and during activation of PV (blue) or SOM (pink) neurons. **f**, **g**, Control tuning curves (black) were suppressed with ChR2 activation of PV neurons (blue) or SOM neurons (pink) in four example cells with very different response levels. **h**, Representative examples of the relationship between PV-mediated (blue) and SOM-mediated (pink) suppression and control response strength. **i**, Bar graph of mean suppression versus response slopes during PV (blue) and SOM (pink) activation. **j**, **k**, Population data for the amount of PV- and SOM-mediated suppression seen at different response levels. **l**, **m**, Average tuning as in Fig. 2l, m, before (black) and during PV (blue) or SOM (pink) activation. **n**–**r**, Effects of PV and SOM activation on tuning-curve parameters, as in Fig. 2n–r. †$P < 0.10$; *$P < 0.05$; **$P < 0.01$; NS, not significant. Data are shown as mean ± s.e.m.

PV-mediated inhibition are distinct, whether they are evoked by populations or single SOM or PV neurons. Furthermore, these maps of affected neurons show remarkable diversity in the functional suppression exerted by specific PV or SOM neurons within their local neighbourhoods.

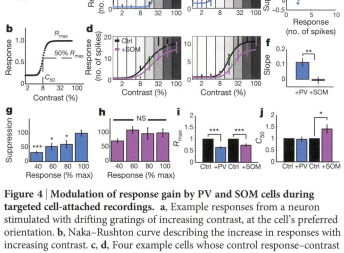


Figure 4 | Modulation of response gain by PV and SOM cells during targeted cell-attached recordings. **a**, Example responses from a neuron stimulated with drifting gratings of increasing contrast, at the cell's preferred orientation. **b**, Naka–Rushton curve describing the increase in responses with increasing contrast. **c**, **d**, Four example cells whose control response–contrast curves (black) were suppressed by PV (blue) or SOM (pink) activation. C_{50} values are marked (green circles). **e**, Relationship between suppression and response when a PV cell was activated (slope = 0.15). **f**, The mean suppression versus response slope was larger when PV cells were activated than when SOM cells were activated. **g**, **h**, Population data for the amount of PV- and SOM-mediated suppression observed at different response levels. **i**, **j**, Effects of PV and SOM activation on response–contrast function parameters R_{max} and C_{50} (see **b**). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS, not significant. Data are shown as mean ± s.e.m.

Functional connectivity of inhibitory networks

An important question is whether there is an underlying logic through which an inhibitory cell makes functional connections with its target cells. We tested whether, *in vivo*, functional suppression by single PV or SOM neurons can be predicted by spatial[22,23] or functional[28,29] relationships. The distance to potential targets did not predict whether a neuron was significantly ($P < 0.05$) suppressed by PV or SOM neurons within 100 μm (Supplementary Fig. 21a), and there was no significant relationship between suppression strength and distance for the population of either PV- or SOM-stimulated networks (Fig. 6a; PV, $P = 0.80$, $R = -0.03$, 4 networks; SOM, $P = 0.1$, $R = -0.28$, 5 networks). Furthermore, the spatial patterns of suppression by PV and SOM neurons were less clustered than a theoretical ordered distance model, and statistically similar to randomly targeted networks, although PV networks tended to be more spatially coherent than SOM networks (Supplementary Fig. 21b–d). Thus, individual PV and SOM neurons do not seem to functionally affect neurons along a distance gradient within their local network.

We then asked whether there was any relationship between the preferred orientation of an inhibitory cell and the preferred orientations of its targets. Comparing the tuning curves of activated inhibitory neurons and the tuning of significantly suppressed cells (Fig. 6b, c), showed that PV cells targeted higher percentages of neurons that matched their own preferred orientations than the orthogonal or the expected percentage predicted by random targeting (Fig. 6d, top panel; preferred orientation (PO), 44.3 ± 7.6%; orthogonal, non-preferred orientation (nonPO), 20.4 ± 4.8%; $P < 0.05$, treating each network as a single observation; $n = 210$ target cells from 4 PV networks; PO versus random PO, $P < 0.05$). However, for SOM cells and networks, the orientation distribution of suppressed cells was more uniform, with no significant difference between the percentage of targeted cells at the preferred orientation versus the orthogonal of the source SOM cells (Fig. 6d, bottom panel; PO, 13.5 ± 3.8%; nonPO, 8.1 ± 4.6%; $P = 0.39$, $n = 238$ target cells from 5 SOM networks). These results indicate that PV cells preferentially target other neurons that have similar preferred

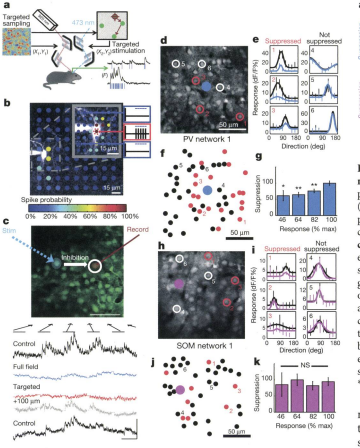


Figure 5 | Dual-laser optical mapping of network connections to reveal maps of functional inhibition by single PV and SOM neurons. a, Two laser systems are controlled independently and synchronized for concurrent high-speed imaging and targeted spatial stimulation of a ChR2-expressing inhibitory neuron. **b**, A SOM ChR2+ neuron *in vivo* was patched and recorded while systematically mapping target locations with a focused 473-nm laser; spikes were elicited only when stimulation was targeted on or very near to the soma. The colour bar indicates the spike probability at each location (0–100%). **c**, All-optical circuit mapping. Optical activation of one neuron (blue spot) during targeted recording of another neuron (red spot). Activity in the recorded neuron was measured with no stimulation (Control, black), full-field activation of PV neurons (blue), targeted activation of the PV neuron (red), and while aiming the stimulation beam 100 μm off-target (grey). Visual responses were suppressed by both full-field and single PV-cell activation. **d**, Mapping the influence of targeted PV activation on neighbouring cells revealed some cells that were significantly suppressed by PV activation (red, 1–3) which were clearly intermixed with other nearby cells that were not affected (white, 4–6). **e**, Tuning curves for cells 1–6 comparing cell responses during control (black) or targeted PV-cell activation (blue). **f**, Spatial distribution of all cells in the network that were either significantly suppressed (red) or not (black). **g**, Population data for the amount of PV- and SOM-mediated suppression observed at cells' different response levels across all single cell networks. **h**–**k**, Same as **d**–**g**, in experiments in which focal stimulation was targeted to single SOM neurons. $*P < 0.05$; $**P < 0.01$; NS, not significant. Data are shown as mean ± s.e.m.

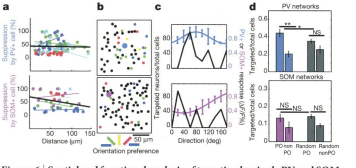


Figure 6 | Spatial and functional analysis of targeting by single PV and SOM neurons. a, The amount of suppression in significantly suppressed cells is plotted against their distance from the stimulated PV cell (top) or SOM cell (bottom). Best fit coloured lines show individual networks; black line depicts pooled data. **b**, The orientation preferences of significantly suppressed cells are colour-coded for two example networks (top, PV (blue); bottom, SOM (pink)). **c**, Distribution of preferred orientations of target cells (black lines) for these example networks, superimposed with the orientation tuning curve of the stimulated PV (blue line, top) or SOM (pink line, bottom) neuron. **d**, Bar graphs show the mean proportion of neurons that matched the preferred orientation of stimulated PV neurons (blue bars) or SOM neurons (pink bars), and the orthogonal, non-preferred orientation. Grey bars show the same comparison when the preferred orientations were randomly re-sampled among the neurons in the fields of view (Random PO and Random nonPO). For PV but not SOM, the actual percentage of targets at the PO was greater than expected with random targeting, indicating that PV cells may preferentially suppress cells with similar functional response properties. $*P < 0.05$; $**P < 0.01$; NS, not significant. Data are shown as mean ± s.e.m.

orientations, whereas SOM neurons seem to have a broader range of targets.

Discussion

Although a growing literature has started to examine the input and firing properties of specific inhibitory neuron classes[19,20,30,31], little is understood about the functional nature of their output. By triggering inhibition and measuring its effects on connected cells in the functioning cortex, we have shown the computational impact of different forms of inhibition on sensory processing. PV neurons principally implement divisive normalization, whereas SOM neurons perform relatively uniform subtraction of responses in their targets, leading to complementary effects on neuronal responses: SOM neurons alter stimulus selectivity, whereas PV neurons preserve selectivity, and PV neurons modulate response gain, whereas SOM neurons shift response levels, leaving response gain unaffected. These effects are mediated by complementary local circuits: PV neurons preferentially target iso-oriented neurons, whereas SOM neurons target cells with a wide range of orientation preferences.

The role of inhibition in shaping stimulus selectivity of visual cortex neurons has been difficult to resolve with previous methods. Intracellular recordings from V1 neurons have found closely matched tuning of excitatory and inhibitory synaptic conductances[8,28,29,32], indicating that inhibition might not sharpen orientation selectivity, and intracellular blockade of inhibition indeed does not seem to affect orientation tuning[33]. However, broadly tuned or untuned inhibition can in principle sharpen neuronal responses; such inhibition has also been described in V1 neurons[10,34,35], and pharmacological blockade of network inhibition broadens orientation selectivity[5–7]. We show that SOM-mediated inhibition targets cells with a range of preferred orientations, and its presence on dendrites probably serves to sharpen the wide orientation preference of dendritic excitatory inputs[36]. Consistent with SOM neurons having this role, genetic reduction of a subset of dendrite-targeting interneurons broadens orientation selectivity[37].

Inhibition has long been proposed to regulate the gain of cortical responses, and we now show that PV cells are crucial for this function. Consistent with our findings, a recent study[26] has demonstrated that activating PV-interneuron populations in mouse visual cortex has a divisive scaling effect on responses of target neurons. Rapid PV-mediated inhibition matched to excitation in time could shape response gain as well as selectivity, as shown in the auditory[9,38], somatosensory[39,40] and prefrontal cortex[41]. Finally, the regulation of response gain by PV cells makes them an attractive mechanism for the developmental regulation of inputs during experience-dependent plasticity of cortical circuits[42].

The distinctive effects of PV and SOM inhibition may arise from their cellular regions of impact[43], and possibly synaptic differences

between the cell types[44,45]. The methods we have described provide a basis for 'functional connectomics' in active cortical circuits, and also reveal the complementary computational roles of specific inhibitory cell classes *in vivo* during sensory processing. In bridging the gap between cellular and network function, these methods should be applicable to many cortical areas and cell types, to elucidate their functional connectivity and embodied computational principles.

METHODS SUMMARY

Details of mice and viral constructs used, animal surgical preparations, *in vitro* slice characterization of ChR2 function, interneuron expression, *in vivo* two-photon guided cell-attached recording, development of high-speed targeted scanning of calcium responses, development of simultaneous two-photon imaging and optogenetic stimulation, the focal stimulation system, and details of data analysis are described in the Supplementary Methods.

Received 24 December 2011; accepted 25 June 2012.
Published online 8 August 2012.

1. Freund, T. F. & Buzsáki, G. Interneurons of the hippocampus. *Hippocampus* **6,** 347–470 (1996).
2. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5,** 793–807 (2004).
3. Klausberger, T. & Somogyi, P. Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* **321,** 53–57 (2008).
4. Lewis, D. A. GABAergic local circuit neurons and prefrontal cortical dysfunction in schizophrenia. *Brain Res. Brain Res. Rev.* **31,** 270–276 (2000).
5. Sillito, A. M. The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *J. Physiol. (Lond.)* **250,** 305–329 (1975).
6. Crook, J. M., Kisvárday, Z. F. & Eysel, U. T. Evidence for a contribution of lateral inhibition to orientation tuning and direction selectivity in cat visual cortex: reversible inactivation of functionally characterized sites combined with neuroanatomical tracing techniques. *Eur. J. Neurosci.* **10,** 2056–2075 (1998).
7. Tsumoto, T., Eckart, W. & Creutzfeldt, O. D. Modification of orientation sensitivity of cat visual cortex neurons by removal of GABA-mediated inhibition. *Exp. Brain Res.* **34,** 351–363 (1979).
8. Ferster, D. & Miller, K. D. Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.* **23,** 441–471 (2000).
9. Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426,** 442–446 (2003).
10. Monier, C., Chavane, F., Baudot, P., Graham, L. J. & Frégnac, Y. Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. *Neuron* **37,** 663–680 (2003).
11. Carandini, M. & Heeger, D. J. Summation and division by neurons in primate visual cortex. *Science* **264,** 1333–1336 (1994).
12. Somers, D. C., Nelson, S. B. & Sur, M. An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* **15,** 5448–5465 (1995).
13. Reynolds, J. H. & Heeger, D. J. The normalization model of attention. *Neuron* **61,** 168–185 (2009).
14. Ohshiro, T., Angelaki, D. E. & DeAngelis, G. C. A normalization model of multisensory integration. *Nature Neurosci.* **14,** 775–782 (2011).
15. Louie, K. & Glimcher, P. W. Separating value from choice: delay discounting activity in the lateral intraparietal area. *J. Neurosci.* **30,** 5498–5507 (2010).
16. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71,** 45–61 (2011).
17. Xu, X., Roby, K. D. & Callaway, E. M. Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *J. Comp. Neurol.* **518,** 389–404 (2010).
18. Schummers, J., Yu, H. & Sur, M. Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex. *Science* **320,** 1638–1643 (2008).
19. Runyan, C. A. *et al.* Response features of parvalbumin-expressing interneurons suggest precise roles for subtypes of inhibition in visual cortex. *Neuron* **67,** 847–857 (2010).
20. Bock, D. D. *et al.* Network anatomy and *in vivo* physiology of visual cortical neurons. *Nature* **471,** 177–182 (2011).
21. Yoshimura, Y. & Callaway, E. M. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nature Neurosci.* **8,** 1552–1559 (2005).
22. Fino, E. & Yuste, R. Dense inhibitory connectivity in neocortex. *Neuron* **69,** 1188–1203 (2011).
23. Packer, A. M. & Yuste, R. Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *J. Neurosci.* **31,** 13260–13271 (2011).
24. Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473,** 87–91 (2011).
25. Lillis, K. P., Eng, A., White, J. A. & Mertz, J. Two-photon imaging of spatially extended neuronal network dynamics with high temporal resolution. *J. Neurosci. Methods* **172,** 178–184 (2008).
26. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73,** 159–170 (2012).
27. Katzner, S., Busse, L. & Carandini, M. GABAA inhibition controls response gain in visual cortex. *J. Neurosci.* **31,** 5931–5941 (2011).
28. Anderson, J. S., Carandini, M. & Ferster, D. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.* **84,** 909–926 (2000).
29. Mariño, J. *et al.* Invariant computations in local cortical networks with balanced excitation and inhibition. *Nature Neurosci.* **8,** 194–201 (2005).
30. Kerlin, A. M., Andermann, M. L., Berezovskii, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67,** 858–871 (2010).
31. Hofer, S. B. *et al.* Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neurosci.* **14,** 1045–1052 (2011).
32. Tan, A. Y., Brown, B. D., Scholl, B., Mohanty, D. & Priebe, N. J. Orientation selectivity of synaptic input to neurons in mouse and cat primary visual cortex. *J. Neurosci.* **31,** 12339–12350 (2011).
33. Nelson, S., Toth, L., Sheth, B. & Sur, M. Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science* **265,** 774–777 (1994).
34. Xing, D., Ringach, D. L., Hawken, M. J. & Shapley, R. M. Untuned suppression makes a major contribution to the enhancement of orientation selectivity in macaque v1. *J. Neurosci.* **31,** 15972–15982 (2011).
35. Liu, B. H. *et al.* Broad inhibition sharpens orientation selectivity by expanding input dynamic range in mouse simple cells. *Neuron* **71,** 542–554 (2011).
36. Jia, H., Rochefort, N. L., Chen, X. & Konnerth, A. Dendritic organization of sensory input to cortical neurons *in vivo*. *Nature* **464,** 1307–1312 (2010).
37. Mao, R. *et al.* Influence of a subtype of inhibitory interneuron on stimulus-specific responses in visual cortex. *Cereb. Cortex* **22,** 493–508 (2011).
38. Tan, A. Y., Zhang, L. I., Merzenich, M. M. & Schreiner, C. E. Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. *J. Neurophysiol.* **92,** 630–643 (2004).
39. Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neurosci.* **11,** 535–537 (2008).
40. Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F. & Petersen, C. C. Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* **65,** 422–435 (2010).
41. Haider, B., Duque, A., Hasenstaub, A. R. & McCormick, D. A. Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* **26,** 4535–4545 (2006).
42. Yazaki-Sugiyama, Y., Kang, S., Câteau, H., Fukai, T. & Hensch, T. K. Bidirectional plasticity in fast-spiking GABA circuits by visual experience. *Nature* **462,** 218–221 (2009).
43. Koch, C., Poggio, T. & Torre, V. Nonlinear interactions in a dendritic tree: localization, timing, and role in information processing. *Proc. Natl Acad. Sci. USA* **80,** 2799–2802 (1983).
44. Kulik, A. *et al.* Compartment-dependent colocalization of Kir3.2-containing K^+ channels and GABAB receptors in hippocampal pyramidal cells. *J. Neurosci.* **26,** 4289–4297 (2006).
45. Nusser, Z., Sieghart, W., Benke, D., Fritschy, J. M. & Somogyi, P. Differential synaptic localization of two major gamma-aminobutyric acid type A receptor alpha subunits on hippocampal pyramidal cells. *Proc. Natl Acad. Sci. USA* **93,** 11939–11944 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Huang for providing the SOM–Cre mouse line; C. Le for performing animal care support and viral injections; S. Yan and Y. Deng for help in the development of optogenetics and imaging methods *in vitro*; S. El-Boustani for collecting data for *in vivo* deconvolution; J. Sharma, M. Goard and A. Banerjee for comments and discussions on the manuscript; L.-H. Tsai, K. Meletis and M. Carlen for early provision of viral constructs and PV–Cre viral injections; and James Schummers and Hiroki Sugihara for participating in early pilot experiments testing optogenetics stimulation *in vivo*. This work was supported by postdoctoral fellowships from the US National Institutes of Health (NIH) and the Simons Foundation (N.R.W.), an NIH predoctoral fellowship (C.A.R.) and grants from the NIH and the Simons Foundation (M.S.).

Author Contributions N.R.W. conceived experiments, designed and engineered circuit interface and analysis systems, carried out *in vivo* and *in vitro* experiments, and performed analyses. C.A.R. conceived experiments, performed surgeries and viral injections, carried out *in vivo* experiments, and performed analyses. F.L.W. carried out *in vivo* experiments, and performed analyses. M.S. conceived experiments and contributed to analysis of experiments. N.R.W., C.A.R. and M.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (msur@mit.edu) or N.R.W. (nathan1@mit.edu).

LETTER

doi:10.1038/nature11379

A massive, cooling-flow-induced starburst in the core of a luminous cluster of galaxies

M. McDonald[1], M. Bayliss[2], B. A. Benson[3,4], R. J. Foley[5], J. Ruel[2], P. Sullivan[1], S. Veilleux[6,7], K. A. Aird[8], M. L. N. Ashby[5], M. Bautz[1], G. Bazin[9,10], L. E. Bleem[3,11], M. Brodwin[12], J. E. Carlstrom[3,4,11,13,14], C. L. Chang[3,4,14], H. M. Cho[15], A. Clocchiatti[16], T. M. Crawford[3,13], A. T. Crites[3,13], T. de Haan[17], S. Desai[9,10], M. A. Dobbs[17], J. P. Dudley[17], E. Egami[18], W. R. Forman[5], G. P. Garmire[19], E. M. George[20], M. D. Gladders[3,13], A. H. Gonzalez[21], N. W. Halverson[22], N. L. Harrington[20], F. W. High[3,13], G. P. Holder[17], W. L. Holzapfel[20], S. Hoover[3,4], J. D. Hrubes[8], C. Jones[5], M. Joy[23], R. Keisler[3,11], L. Knox[24], A. T. Lee[19,25], E. M. Leitch[3,13], J. Liu[9,10], M. Lueker[19,26], D. Luong-Van[8], A. Mantz[3], D. P. Marrone[18], J. J. McMahon[3,4,27], J. Mehl[3,13], S. S. Meyer[3,4,11,13], E. D. Miller[1], L. Mocanu[3,13], J. J. Mohr[9,10,28], T. E. Montroy[29], S. S. Murray[5], T. Natoli[3,11], S. Padin[3,13,26], T. Plagge[3,13], C. Pryke[30], T. D. Rawle[18], C. L. Reichardt[20], A. Rest[31], M. Rex[18], J. E. Ruhl[29], B. R. Saliwanchik[29], A. Saro[9,10], J. T. Sayre[29], K. K. Schaffer[3,4,32], L. Shaw[17,33], E. Shirokoff[20,25], R. Simcoe[1], J. Song[27], H. G. Spieler[25], B. Stalder[5], Z. Staniszewski[29], A. A. Stark[5], K. Story[3,11], C. W. Stubbs[2,5], R. Šuhada[9], A. van Engelen[17], K. Vanderlinde[17], J. D. Vieira[3,11,26], A. Vikhlinin[5], R. Williamson[3,13], O. Zahn[20,34,35] & A. Zenteno[9,10]

In the cores of some clusters of galaxies the hot intracluster plasma is dense enough that it should cool radiatively in the cluster's lifetime[1–3], leading to continuous 'cooling flows' of gas sinking towards the cluster centre, yet no such cooling flow has been observed. The low observed star-formation rates[4,5] and cool gas masses[6] for these 'cool-core' clusters suggest that much of the cooling must be offset by feedback to prevent the formation of a runaway cooling flow[7–10]. Here we report X-ray, optical and infrared observations of the galaxy cluster SPT-CLJ2344-4243 (ref. 11) at redshift $z = 0.596$. These observations reveal an exceptionally luminous (8.2×10^{45} erg s^{-1}) galaxy cluster that hosts an extremely strong cooling flow (around 3,820 solar masses a year). Further, the central galaxy in this cluster appears to be experiencing a massive starburst (formation of around 740 solar masses a year), which suggests that the feedback source responsible for preventing runaway cooling in nearby cool-core clusters may not yet be fully established in SPT-CLJ2344-4243. This large star-formation rate implies that a significant fraction of the stars in the central galaxy of this cluster may form through accretion of the intracluster medium, rather than (as is currently thought) assembling entirely via mergers.

The galaxy cluster SPT-CLJ2344-4243 was discovered by the South Pole Telescope[12] via the Sunyaev–Zel'dovich effect, with an initial estimated[11] mass of $16.6 \times 10^{14} M_{\odot}$, where $M_{\odot}$ is the mass of the Sun. These data were supplemented with new broadband optical *g*, *r*, *i*, *z* imaging from the Mosaic II camera on the Blanco 4-m telescope (Fig. 1), optical spectroscopy from the Gemini Multi-Object Spectrograph (GMOS) on the Gemini South 8.1-m telescope, optical long-slit spectroscopy using the Inamori-Magellan Areal Camera and Spectrograph (IMACS) on the 6.5-m Magellan telescope, near-infrared long-slit spectroscopy using the Folded-port InfraRed Echellette (FIRE) on the 6.5-m Magellan telescope, mid-far-infrared imaging using the Photodetector Array Camera and Spectrometer (PACS) and the Spectral and Photometric Imaging Receiver (SPIRE) on the Herschel Space Observatory, and X-ray imaging spectroscopy using the ACIS-I camera on the Chandra X-ray Observatory. Additionally, we have acquired archival near–far-ultraviolet imaging from the orbiting space telescope Galaxy Evolution Explorer (GALEX) archives, near–mid-infrared imaging from the Two Micron All-Sky Survey (2MASS) and Wide-field Infrared Survey Explorer (WISE) archives, and 843-MHz radio imaging from the Sydney University Molonglo Sky Survey (SUMSS) survey. Further details of these data and their processing can be found in the Supplementary Information.

We estimate the mass of SPT-CLJ2344-4243 from the X-ray-measured pressure ($Y_X \equiv M_{\rm gas} \times T_X$, where T_X is the X-ray-measured temperature of the intracluster gas and $M_{\rm gas}$ is its mass) of the intracluster medium, using an externally calibrated pressure–mass (Y_X–M) relation. The relation was calibrated using a local sample of relaxed (that is, non-merging) clusters from X-ray estimates of the total mass that assumed hydrostatic equilibrium[13]. By iteratively adjusting the value of r_{500} (where r_{500} (or r_{200}) is the radius for which the enclosed average density is 500 (or 200) times the critical (or average) density of the Universe) such that the Y_X–M_{500} relation is satisfied, we converge on values of $r_{500} = 1.3$ Mpc

[1]MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [2]Department of Physics, Harvard University, 17 Oxford Street, Cambridge, Massachusetts 02138, USA. [3]Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. [4]Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. [5]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [6]Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. [7]Astroparticle Physics Laboratory, NASA Goddard Space Flight Center, Code 661, Greenbelt, Maryland 20771, USA. [8]University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. [9]Department of Physics, Ludwig-Maximilians-Universität, Scheinerstrasse 1, 81679 München, Germany. [10]Excellence Cluster Universe, Boltzmannstrasse 2, 85748 Garching, Germany. [11]Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. [12]Department of Physics and Astronomy, University of Missouri, 5110 Rockhill Road, Kansas City, Missouri 64110, USA. [13]Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. [14]Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. [15]NIST Quantum Devices Group, 325 Broadway Mailcode 817.03, Boulder, Colorado 80305, USA. [16]Departamento de Astronomia y Astrofisica, Pontificia Universidad Catolica, Santiago, Chile. [17]Department of Physics, McGill University, 3600 Rue University, Montreal, Quebec H3A 2T8, Canada. [18]Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. [19]Department of Astronomy and Astrophysics, Pennsylvania State University, 525 Davey Laboratory, University Park, Pennsylvania 16802, USA. [20]Department of Physics, University of California, Berkeley, California 94720, USA. [21]Department of Astronomy, University of Florida, Gainesville, Florida 32611, USA. [22]Department of Astrophysical and Planetary Sciences and Department of Physics, University of Colorado, Boulder, Colorado 80309, USA. [23]Department of Space Science, VP62, NASA Marshall Space Flight Center, Huntsville, Alabama 35812, USA. [24]Department of Physics, University of California, One Shields Avenue, Davis, California 95616, USA. [25]Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [26]California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA. [27]Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, Michigan 48109, USA. [28]Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 85748 Garching, Germany. [29]Physics Department, Center for Education and Research in Cosmology and Astrophysics, Case Western Reserve University, Cleveland, Ohio 44106, USA. [30]Physics Department, University of Minnesota, 116 Church Street Southeast, Minneapolis, Minnesota 55455, USA. [31]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. [32]Liberal Arts Department, School of the Art Institute of Chicago, 112 South Michigan Avenue, Chicago, Illinois 60603, USA. [33]Department of Physics, Yale University, PO Box 208210, New Haven, Connecticut 06520-8120, USA. [34]Berkeley Center for Cosmological Physics, Department of Physics, University of California, California 94720, USA. [35]Lawrence Berkeley National Laboratories, Berkeley, California 94720, USA.

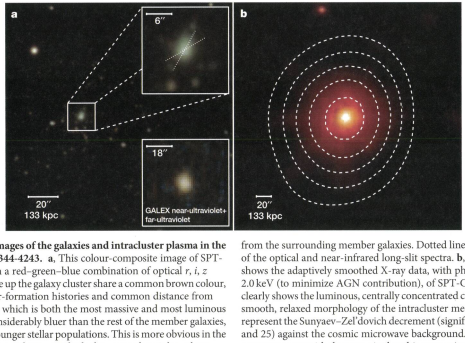


Figure 1 | False-colour images of the galaxies and intracluster plasma in the galaxy cluster SPT-CLJ2344-4243. a, This colour-composite image of SPT-CLJ2344-4243 is based on a red–green–blue combination of optical *r*, *i*, *z* images. Galaxies that make up the galaxy cluster share a common brown colour, owing to their similar star-formation histories and common distance from Earth. The central galaxy, which is both the most massive and most luminous galaxy in the cluster, is considerably bluer than the rest of the member galaxies, suggesting significantly younger stellar populations. This is more obvious in the zoomed-in inset. The lower right inset, which shows an ultraviolet colour-composite, reveals a bright ultraviolet source, with no accompanying emission from the surrounding member galaxies. Dotted lines represent the orientation of the optical and near-infrared long-slit spectra. **b**, This false-colour image shows the adaptively smoothed X-ray data, with photon energies from 0.7–2.0 keV (to minimize AGN contribution), of SPT-CLJ2344-4243. The image clearly shows the luminous, centrally concentrated core, as well as the relatively smooth, relaxed morphology of the intracluster medium. White contours represent the Sunyaev–Zel'dovich decrement (significance levels of 5, 10, 15, 20 and 25) against the cosmic microwave background. The circularity of these contours agree with the scenario that this system is not currently undergoing a major merger with another galaxy cluster.

and $M_{500,\,Y_X} = 12.6^{+2.0}_{-1.5} \times 10^{14} M_{\odot}$. At r_{200}, this corresponds to $M_{200,Y_X} \approx 25 \times 10^{14} M_{\odot}$, which makes SPT-CLJ2344-4243 amongst the most massive known clusters in the Universe[14,15]. The GMOS multi-object spectroscopy of 26 galaxies, each exhibiting only absorption features, yielded a robust biweight estimate of the redshift ($z = 0.596 \pm 0.002$) and velocity dispersion ($1{,}700^{+300}_{-200}$ km s^{-1}), the latter being consistent with the picture of an extremely massive cluster. The velocity distribution is consistent with a Gaussian distribution, but the limited number of redshifts does not preclude velocity substructure or multimodality. The smooth X-ray isophotes suggest that the cluster may be relaxed, and although the cluster member velocity distribution is consistent with an undisturbed cluster, the velocity data lack the statistical power to constrain the cluster's dynamical state robustly.

The integrated rest-frame X-ray luminosity at 2–10 keV, $L_{2-10\,\mathrm{keV}} = 8.2^{+0.1}_{-0.2} \times 10^{45}$ erg s^{-1} within r_{500}, is greater than any other known cluster in this band. The high central luminosity, which is predominantly cooling radiation, in turn results in a high X-ray cooling rate, as defined by $\frac{\mathrm{d}M}{\mathrm{d}t} = \frac{2L\mu m_{\mathrm{p}}}{5kT}$, where μ is the mean molecular mass of the intracluster medium, L is the X-ray luminosity, m_{p} is the mass of a proton and kT is the temperature of the intracluster gas. Assuming a cooling radius of 100 kpc (see Supplementary Information), we measure an intracluster medium cooling rate of $3{,}820 \pm 530 M_{\odot}$ yr^{-1}, making this the strongest cooling flow yet discovered (see Table 1 for comparison to other clusters). The intracluster medium in SPT-CLJ2344-4243 exhibits a significant drop in temperature, accompanied by a rise in the metallicity in the central 100 kpc that is reminiscent of nearby cool core clusters. Furthermore, the short central cooling time (less than a billion years, <1 Gyr), along with the low central entropy (<100 keV cm^2), resembles nearby strongly cooling cores, such as the Perseus[16] and PKS0745-191 (ref. 17) clusters. The discovery of a strong cool core at $z = 0.596$ is particularly remarkable because recent X-ray and optical surveys have found a general lack of strong cool cores at $z > 0.4$ (refs 18–20), with relatively few exceptions.

Much like the central galaxies of low-z cool-core clusters[21–23], SPT-CLJ2344-4243 exhibits bright, spatially extended, optical line emission (that is, [O II], Hβ, [O III], [O I], Hα, [N II], [S II], and so on; Fig. 2). We were fortunate to intersect what appears to be an extended filament with one of our randomly oriented slits, which has a length of about 70 kpc. This is similar in extent to the most extended optical filaments in the core of the Perseus cluster[24], and orders of magnitude larger than typical jets in cluster cores (such as M87; ref. 25). The diagnostic line ratios [N II]/Hα, [S II]/Hα, [O I]/Hα, [O III]/Hβ, and [O II]/[O II] show evidence for a Seyfert-like active galactic nucleus (AGN) in the central galaxy, while at radii exceeding 2″, the optical line ratios resemble those in the star-forming filaments of $z \approx 0$ cool-core clusters[26] (see Supplementary Information).

Apart from the exceptionally high X-ray luminosity and central cooling rate, what sets this system apart from the majority of nearby galaxy clusters is that there is significant evidence for a dusty starburst in the central galaxy of SPT-CLJ2344-4243. The rest-frame 0.1–500 μm spectral energy distribution (Fig. 3) of the central galaxy most closely resembles that of an ultraluminous, infrared galaxy; such galaxies are known to have heavily obscured starbursts (about $200-1{,}000 M_{\odot}$ yr^{-1}) and central AGNs. This scenario is corroborated by our observation of significant Balmer reddening ($E(B-V)_{\mathrm{global}} \approx 0.3$)

Table 1 | Well-studied, strong, cool-core clusters compared to SPT-CLJ2344-4243

Cluster	z	$L_{2.0\text{-}10.0\,\mathrm{keV}}$ (10^{44} erg s^{-1})	kT (keV)	dM/dt ($M_{\odot}$ yr^{-1})	SFR ($M_{\odot}$ yr^{-1})	ϵ_{cool}
Perseus	0.0179	11	5.5	556	37	0.07
PKS0745-191	0.1028	29.5	6.71	1,455	20	0.01
Zw 3146	0.2906	36.9	6.4	2,228	79	0.04
RX J1347.5-1145	0.451	60	10.0	1,900	23	0.01
SPT-CLJ2344-4243	0.596	82^{+1}_{-2}	$13.0^{+2.4}_{-3.4}$	3,820 ± 530	740 ± 160	0.19 ± 0.05

Prior to this work, RX J1347.5-1145 was considered both the most X-ray luminous and strongest cooling galaxy cluster, with a luminosity of $L_{2-10\,\mathrm{keV}} = 60 \times 10^{44}$ erg s^{-1} and cooling rate of 1,900$M_{\odot}$ yr^{-1}. The star-formation rate of the central galaxy in SPT-CLJ2344-4243 is exceptionally high. We quantify the efficiency of converting the cooling flow into stars with the parameter ϵ_{cool}, which is simply the star-formation rate normalized to the classical cooling rate. The high star-formation rate implies that SPT-CLJ2344-4243 is converting around 20% of the cooling flow into stars, which is considerably higher than the vast majority of low-redshift cool-core clusters. X-ray properties and star-formation rates of the lower-redshift clusters are taken from the literature[26,27].

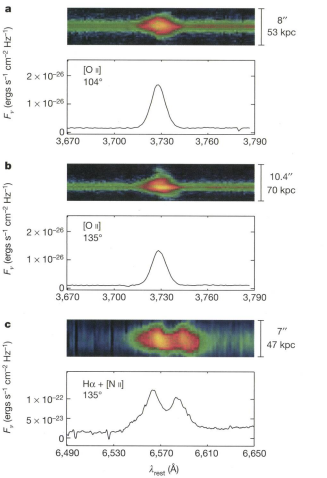


Figure 2 | Optical and near-infrared emission-line spectra of the central galaxy in SPT-CLJ2344-4243. The false-colour images highlight the extended line emission ([O II] in **a** and **b**, Hα + [N II] in **c**), where the vertical axis is the spatial direction (along the slit) and the horizontal axis is the spectral direction. F_ν is the emission flux. These emission lines result from warm, ionized gas at about 10^4 K, which is probably heated by a combination of young stars, shocks and feedback from the central active black hole. The fact that the [O II] emission is significantly more extended in one direction (position angle of the slit, relative to north $\theta \approx 135^\circ$) suggests that the emission is non-axisymmetric, and is consistent with the scenario of radial line-emitting filaments. The extent of this emission (>50 kpc) is reminiscent of optical filaments observed in the core of the Perseus cluster[24]. Beneath each colour image, we show the spectrum that is generated by summing along columns of the colour image. These spectra show the high signal-to-noise ratio of these emission lines, leading to high-confidence estimates of the emission line luminosity. **c**, The near-infrared spectrum (in the observed frame, $\lambda_{H\alpha} \approx 1.05\,\mu m$) shows emission from both the Hα and [N II] lines, extended over similar radii (>50 kpc).

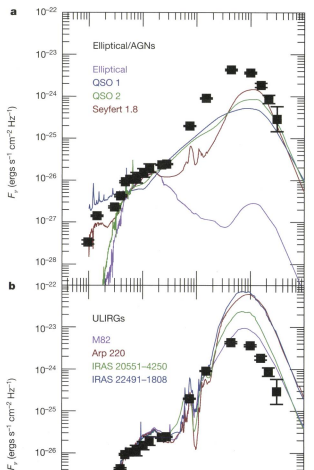


Figure 3 | Spectral energy distribution of the central galaxy in SPT-CLJ2344-4243 from the far-ultraviolet to the far-infrared. We show on the *y*-axis the integrated specific flux (F_ν) as a function of wavelength in the rest frame of the galaxy cluster, along with the associated 1σ measurement error. A typical central cluster galaxy is morphologically classified as an 'elliptical' galaxy, and has an average spectral energy distribution like that shown in purple in **a**. In contrast, the central galaxy in SPT-CLJ2344-4243 has a considerable excess of emission at both ultraviolet and infrared wavelengths, indicative of strong star formation. While there is significant evidence for strong feedback from the central, supermassive black hole in this cluster (see Supplementary Information), this figure shows that simple models of AGNs (quasars type 1 and 2, and Seyfert type 1.8 shown in **a**) are unable to reproduce the spectral shape of the central galaxy in SPT-CLJ2344-4243. However, in **b** we show templates[40] of four different dusty starbursts, or ultra-luminous infrared galaxies (ULIRGs), which provide a much better match to the data. This good agreement suggests that the central galaxy in SPT-CLJ2344-4243, unlike typical central cluster galaxies, contains a dusty starburst and a heavily obscured AGN. Specifically, the spectral shape is most similar to those of M82, a dusty starburst with a strong wind, and IRAS 20551-4250, which is a composite of a highly obscured AGN and a vigorous, dusty starburst.

and strong 24–160 μm emission, combined with signatures of ongoing star formation (near- and far-ultraviolet emission, bright nebular emission lines, weak 4,000-Å break) and a heavily obscured central AGN ($E(B-V)_{nuclear} \approx 0.5$, $n_{H,X\text{-}ray} \approx 40 \times 10^{22}\,cm^{-2}$). Using the full multi-wavelength data set, which includes X-ray, near–far ultraviolet, optical, near–far infrared, and radio data, we estimate an extinction-corrected, AGN-subtracted star-formation rate of $740 \pm 160 M_\odot\ yr^{-1}$, assuming a geometric correction of 45% for the long-slit spectroscopy and an AGN contamination fraction of around 40%–50% (see Supplementary Information for details).

The presence of extended (around 70 kpc), morphologically complex (Fig. 2), star-forming filaments coincident with the central galaxy in SPT-CLJ2344-4243 is reminiscent of low-z cool-core clusters like Perseus and PKS0745-191. However, while these clusters have substantial amounts of star formation (around 20 solar masses a year)[4,5], this is still orders of magnitude less than predicted by the classical cooling estimates based on the X-ray luminosity (Table 1). This disagreement has become known as the 'cooling flow' problem, and it is generally assumed that some form of feedback is responsible for halting the cooling intracluster medium before it reaches the cold

phase. SPT-CLJ2344-4243, however, represents one of a very few exceptions to this general trend[4,27], where the high star-formation rate represents a significant fraction of the massive cooling flow (Table 1). Whatever feedback mechanism is responsible for preventing runaway cooling of the intracluster medium in low-redshift galaxy clusters is clearly operating with a lower efficiency in this system. Although the central galaxy hosts an AGN, as shown by a hard X-ray point source and strong radio emission (see Supplementary Information), it may be that we are observing this system during a small window in time when the AGN is rapidly feeding off the cooling flow, but the power output of the AGN has not yet fully coupled to the intracluster medium, and therefore is able to halt a smaller fraction of the total cooling than in typical low-redshift clusters (such as Perseus). That systems with such high cooling and star-formation rates are not observed at $z = 0$ suggests that either this system is unique, or that the mechanism which quenches cooling may have been less effective in the early Universe. Further studies of distant, strongly cooling galaxy clusters are needed to differentiate between these two scenarios.

The high star-formation rates inferred from optical line emission and near–far-ultraviolet, optical, and mid–far-infrared continuum emission, combined with the strong signatures of X-ray cooling, suggest that the central galaxy in SPT-CLJ2344-4243 may form a substantial fraction of its stars through an intense, short-lived cooling phase of the intracluster medium. Such strong cooling cannot be sustained for a significant amount of time, or both the central galaxy and its supermassive black hole would become too massive, and the central galaxy would have stellar populations considerably younger than those observed in giant elliptical galaxies today. This implies that episodes of strong cooling are short-lived, in contrast to the longer episodes of strong feedback observed in nearby clusters.

Received 5 April; accepted 12 July 2012.

1. Lea, S. M., Silk, J., Kellogg, E. & Murray, S. Thermal-Bremsstrahlung interpretation of cluster X-ray sources. *Astrophys. J.* **184,** L105 (1973).
2. Cowie, L. L. & Binney, J. Radiative regulation of gas flow within clusters of galaxies—a model for cluster X-ray sources. *Astrophys. J.* **215,** 723–732 (1977).
3. Fabian, A. C. & Nulsen, P. E. J. Subsonic accretion of cooling gas in clusters of galaxies. *Mon. Not. R. Astron. Soc.* **180,** 479–484 (1977).
4. O'Dea, C. P. *et al.* An infrared survey of brightest cluster galaxies. II. Why are some brightest cluster galaxies forming stars? *Astrophys. J.* **681,** 1035–1045 (2008).
5. McDonald, M., Veilleux, S., Rupke, D. S. N., Mushotzky, R. & Reynolds, C. Star formation efficiency in the cool cores of galaxy clusters. *Astrophys. J.* **734,** 95 (2011).
6. Edge, A. C. The detection of molecular gas in the central galaxies of cooling flow clusters. *Mon. Not. R. Astron. Soc.* **328,** 762–782 (2001).
7. McNamara, B. R. & Nulsen, P. E. J. Heating hot atmospheres with active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **45,** 117–175 (2007).
8. Fabian, A. C. Observational evidence of AGN feedback. *Annu. Rev. Astron. Astrophys.* (in the press); preprint at http://arxiv.org/abs/1204.4114 (2012).
9. Mathews, W. G. Stopping cooling flows with cosmic-ray feedback. *Astrophys. J. Lett.* **695,** L49–L52 (2009).
10. Gómez, P. L., Loken, C., Roettiger, K. & Burns, J. O. Do cooling flows survive cluster mergers? *Astrophys. J.* **569,** 122–133 (2002).
11. Williamson, R. *et al.* A Sunyaev-Zel'dovich-selected sample of the most massive galaxy clusters in the 2500 deg^2 South Pole Telescope Survey. *Astrophys. J.* **738,** 139 (2011).
12. Carlstrom, J. E. *et al.* The 10 meter South Pole Telescope. *Proc. Astron. Soc. Pacif.* **123,** 568–581 (2011).
13. Vikhlinin, A. *et al.* Chandra Cluster Cosmology Project. II. Samples and X-ray data reduction. *Astrophys. J.* **692,** 1033–1059 (2009).
14. Menanteau, F. *et al.* The Atacama Cosmology Telescope: ACT-CL J0102–4915 "El Gordo", a massive merging cluster at redshift 0.87. *Astrophys. J.* **748,** 7 (2011).
15. Foley, R. J. *et al.* Discovery and cosmological implications of SPT-CL J2106–5844, the most massive known cluster at $z>1$. *Astrophys. J.* **731,** 86 (2011).
16. Fabian, A. C. *et al.* Chandra imaging of the complex X-ray core of the Perseus cluster. *Mon. Not. R. Astron. Soc.* **318,** L65–L68 (2000).
17. Allen, S. W., Fabian, A. C. & Kneib, J. P. A combined X-ray and gravitational lensing study of the massive cooling-flow cluster PKS 0745–191. *Mon. Not. R. Astron. Soc.* **279,** 615–635 (1996).
18. Vikhlinin, A. *et al.* in *Heating versus Cooling in Galaxies and Clusters of Galaxies* (eds Böhringer, H., Pratt, G. W., Finoguenov, A. & Schuecker, P.) 48 (2007); preprint at http://arxiv.org/abs/astro-ph/0611438.
19. Santos, J. S. *et al.* Searching for cool core clusters at high redshift. *Astron. Astrophys.* **483,** 35–47 (2008).
20. McDonald, M. Optical line emission in brightest cluster galaxies at $0 < z 0.6$: evidence for a lack of strong cool cores 3.5 Gyr ago? *Astrophys. J. Lett.* **742,** L35 (2011).
21. Hu, E. M., Cowie, L. L. & Wang, Z. Long-slit spectroscopy of gas in the cores of X-ray luminous clusters. *Astrophys. J.* **59** (Suppl.), 447–498 (1985).
22. Heckman, T. M., Baum, S. A., van Breugel, W. J. M. & McCarthy, P. Dynamical, physical, and chemical properties of emission-line nebulae in cooling flows. *Astrophys. J.* **338,** 48–77 (1989).
23. McDonald, M., Veilleux, S., Rupke, D. S. N. & Mushotzky, R. On the origin of the extended Hα filaments in cooling flow clusters. *Astrophys. J.* **721,** 1262–1283 (2010).
24. Conselice, C. J., Gallagher, J. S. III & Wyse, R. F. G. On the nature of the NGC 1275 system. *Astron. J.* **122,** 2281–2300 (2001).
25. Sparks, W. B., Biretta, J. A. & Macchetto, F. The jet of M87 at tenth-arcsecond resolution: optical, ultraviolet, and radio observations. *Astrophys. J.* **473,** 254 (1996).
26. McDonald, M., Veilleux, S. & Rupke, D. S. N. Optical spectroscopy of Hα filaments in cool core clusters: kinematics, reddening, and sources of ionization. *Astrophys. J.* **746,** 153 (2012).
27. McNamara, B. R. *et al.* The starburst in the Abell 1835 Cluster Central Galaxy: a case study of galaxy formation regulated by an outburst from a supermassive black hole. *Astrophys. J.* **648,** 164–175 (2006).
28. Allen, S. W. The properties of cooling flows in X-ray luminous clusters of galaxies. *Mon. Not. R. Astron. Soc.* **315,** 269–295 (2000).
29. Gitti, M. & Schindler, S. XMM-Newton observation of the most X-ray-luminous galaxy cluster RX J1347.5–1145. *Astron. Astrophys.* **427,** L9–L12 (2004).
30. Poletta, M. *et al.* Spectral energy distributions of hard X-ray selected active galactic nuclei in the XMM-Newton Medium Deep Survey. *Astrophys. J.* **663,** 81–102 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements M.McD. was supported at MIT by NASA through the Chandra X-ray Observatory. The South Pole Telescope is supported by the National Science Foundation, with partial support provided by the Kavli Foundation, and the Moore Foundation. Support for X-ray analysis was provided by NASA. Work at McGill University is supported by NSERC, the CRC programme, and CIfAR, and at Harvard University by the NSF. S.V. acknowledges a Senior NPP Award held at the NASA Goddard Space Flight Center. R.K. acknowledges a NASA Hubble Fellowship, B.A.B. acknowledges a KICP Fellowship, M.A.D. acknowledges an Alfred P. Sloan Research Fellowship, and O.Z. acknowledges a BCCP fellowship.

Author Contributions M.McD. reduced the X-ray and optical long slit spectroscopic data, performed the main analysis, and wrote the paper, with significant assistance from B.A.B., R.J.F. and S.V., and comments from all other authors. M.B. and J.R. reduced multi-slit observations of SPT-CL2344-4243 and performed the velocity dispersion analysis. P.S. and R.S. obtained the infrared spectroscopy, and P.S. reduced these data. All other authors (listed alphabetically) have contributed as part of the South Pole Telescope collaboration, by either their involvement with the initial cluster discovery with the South Pole Telescope and/or multi-wavelength follow-up.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.McD. (mcdonald@space.mit.edu).

LETTER

doi:10.1038/nature11339

Room-temperature solid-state maser

Mark Oxborrow[1], Jonathan D. Breeze[2] & Neil M. Alford[2]

The invention of the laser has resulted in many innovations, and the device has become ubiquitous. However, the maser, which amplifies microwave radiation rather than visible light, has not had as large an impact, despite being instrumental in the laser's birth[1,2]. The maser's relative obscurity has mainly been due to the inconvenience of the operating conditions needed for its various realizations: atomic[3] and free-electron[4] masers require vacuum chambers and pumping; and solid-state masers[5], although they excel as low-noise amplifiers[6] and are occasionally incorporated in ultrastable oscillators[7,8], typically require cryogenic refrigeration. Most realizations of masers also require strong magnets, magnetic shielding or both. Overcoming these various obstacles would pave the way for improvements such as more-sensitive chemical assays, more-precise determinations of biomolecular structure and function, and more-accurate medical diagnostics (including tomography) based on enhanced magnetic resonance spectrometers[9] incorporating maser amplifiers and oscillators. Here we report the experimental demonstration of a solid-state maser operating at room temperature in pulsed mode. It works on a laboratory bench, in air, in the terrestrial magnetic field and amplifies at around 1.45 gigahertz. In contrast to the cryogenic ruby maser[6], in our maser the gain medium is an organic mixed molecular crystal, *p*-terphenyl doped with pentacene, the latter being photo-excited by yellow light. The maser's pumping mechanism exploits spin-selective molecular intersystem crossing[10] into pentacene's triplet ground state[11,12]. When configured as an oscillator, the solid-state maser's measured output power of around −10 decibel milliwatts is approximately 100 million times greater than that of an atomic hydrogen maser[3], which oscillates at a similar frequency (about 1.42 gigahertz). By exploiting the high levels of spin polarization readily generated by intersystem crossing in photo-excited pentacene and other aromatic molecules, this new type of maser seems to be capable of amplifying with a residual noise temperature far below room temperature.

To the best of our knowledge, no-one has previously attempted to make a zero-field maser[13] out of an optically pumped crystal of pentacene-doped *p*-terphenyl. Two further novel features of the device we report here are a microwave cavity containing a dielectric ring of sapphire, which supports a high-Q electromagnetic mode (where Q is the mode's quality factor), and a medical laser, designed for the treatment of vascular skin lesions, that is here used to deliver intense pulses of pump light into the pentacene:*p*-terphenyl maser crystal at its optimal absorption wavelength (~585 nm).

Because they are key components within many measurement and telecommunication devices and systems, there has been intensive research on low-noise amplifiers for the purpose of improving their performance and reducing their operating requirements[14]. As low-noise amplifiers, conventional solid-state masers[5] offer low residual noise temperatures[6] and, in contrast to microfabricated devices, are extremely rugged with respect to electromagnetic shock and overloading. Furthermore, they have low intermodulation distortion[6], so reducing interference from out-of-band signals, and low flicker noise[7], thus enabling precise measurements. The active material within a conventional solid-state maser is a dielectric crystal containing paramagnetic ions, for example ruby, that is maintained typically below 10 K and exposed (with exceptions[8,13]) to a strong d.c. magnetic bias field. Despite the inconvenience of these operating conditions, as well as competition from amplifiers based on cooled high-electron-mobility transistors (HEMTs)[15], ruby masers are used in deep-space communications[6]. They were used, for example, by NASA to receive images of the Solar System's outer planets and their moons from the Voyager space probes[16].

The lowest-noise semiconductor-based amplifiers at microwave frequencies, which today most often incorporate cooled HEMTs[15], suffer from a particular limitation. On cooling, the HEMT amplifier's residual noise temperature, although low enough for many applications, is difficult to reduce below a few kelvin[17]. Although this 'HEMT plateau' can be lowered somewhat by using lower device powers[18], it has motivated the successful development of amplifiers based on resonant superconducting interference devices[17], which offer substantially lower noise temperatures, albeit over more limited bandwidths. But with both HEMTs and such devices, the attainment of millikelvin noise temperatures demands subkelvin refrigeration, as supplied by a dilution refrigerator or its equivalent.

It is known why conventional solid-state masers[5,6] cannot be made to work at room temperature[19]. The most serious problem is that the rate of spin–lattice relaxation, $1/T_1$, and, thus, the microwave power required to saturate the maser's pump transition and, in turn, the required thermal cooling power of the maser's refrigerator, increase extremely rapidly with the absolute (lattice) temperature, T, of the maser crystal. For two-phonon (Raman) scattering off Kramers-type paramagnetic ions in a three-dimensional lattice well below its Debye temperature, this relaxation rate scales as T^9. A second problem is that, even if the maser's pump transition can be saturated by pumping hard, the population inversion and, hence, the maser crystal's gain scale as $I \times hf_{\text{mas.}}/kT$, where $I \equiv f_{\text{pump}}/2f_{\text{mas.}} - 1$ is the inversion ratio[5], $f_{\text{mas.}}$

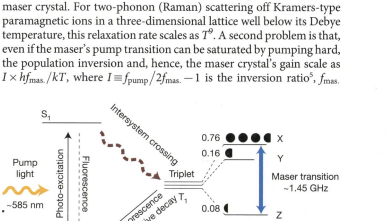


Figure 1 | Maser pumping scheme (Jablonski diagram). Pentacene guest molecules within a *p*-terphenyl host lattice are driven from their singlet ground states, S_0, into their first-excited single states, S_1, by absorbing photons of yellow pump light. From S_1, they predominantly undergo intersystem crossing into their triplet ground states, T_1. This process is spin selective; the relative population rates[11,12] (splitting ratios) into the three spin sublevels of T_1 are as stated (also represented by solid black circles). As can be seen, the uppermost sublevel, X, is preferentially populated, resulting in a strong initial population inversion between it and the lowest sublevel, Z, so providing the conditions for masing through the $X \leftrightarrow Z$ transition.

[1]National Physical Laboratory, Hampton Road, Teddington TW11 0LW, UK. [2]Imperial College, Department of Materials, South Kensington Campus, London SW7 2AZ, UK.

and f_{pump} are respectively the maser signal and pump frequencies, and k and h are respectively Boltzmann's and Planck's constants. The residual noise temperature of a conventionally pumped solid-state maser amplifier thus scales as T, so nullifying the maser's low-noise advantage at higher operating temperatures. It was realized very early on that, at the expense of energy efficiency, larger population inversions and, thus, lower noise temperatures and somewhat higher sustainable operating temperatures can be obtained, for a fixed f_{max}, by increasing f_{pump}—in the extreme case by driving optical pump transitions with light[20], which can often be supplied by a laser. But optical pumping alone does not obviate the first problem, that is, the rapid increase in the rate of spin–lattice relaxation with temperature[21].

Steps have been taken towards the realization of solid-state masers at room temperature using optically excited organic molecules in strong d.c. magnetic fields, focusing largely on C_{60} and porphyrins known to have long spin-polarization lifetimes[19] (T_1). But the 'triplet mechanism', of which Fig. 1 is an example, was considered unsuitable for maser operation at room temperature[22]. Maser threshold was almost reached by instead using the more complex, indirect, radical-triplet-pair mechanism (the triplet state in question being that of etioporphyrin); see fig. 3 of ref. 23. More recently, the potential of various triplet-mechanism photochemistries, including the fullerene derivative PCBM in polystyrene, was assessed in the context of realizing particle acceleration by stimulated emission of radiation at X-band microwave frequencies[24].

The maser reported here (Figs 1 and 2) is based on the triplet mechanism. It exploits spin-selective intersystem crossing[10] into the triplet ground state of pentacene molecules doping crystalline *p*-terphenyl[11], as has been studied in connection with realizations of both dynamic nuclear polarization[25] and electron spin resonance of single molecules[12]. On photo-excitation from pentacene's ground state, S_0, to its first excited singlet state, S_1, with yellow pump light (Fig. 1), intersystem crossing provides a strongly inverted spin population between the uppermost (X) and lowest (Z) spin sublevels of pentacene's triplet ground state[11,12]. The relatively high zero-field splitting (as compared with PCBM, for example) makes it practicable to use a microwave resonator with no applied d.c. magnetic field, which greatly simplifies the experiment set-up. Our chosen design of resonator (Fig. 2) is cylindrical and supports a frequency-tunable transverse-electric mode, $TE_{01\delta}$, which has no azimuthal dependence, has a single antinode in the radial and the axial directions, and has a mode volume of $\sim$50 cm^3. With the pentacene:*p*-terphenyl crystal in place, this mode's unloaded quality factor (Q) was measured to be around 2×10^5. When critically coupled through a magnetic loop probe in the side of the resonator's enclosing metal can (Fig. 2), the mode offered a field-to-power conversion factor of $A \approx 5\,\mathrm{G\,W^{-1/2}}$, which is proportional to the square root of the mode's magnetic Purcell factor[26].

An unpolarized, $\sim$10-mm-diameter beam of yellow (585 $\pm$ 5 nm) pump light from a Candela SPTL-1b long-pulsed rhodamine-6G dye laser was directed at the maser crystal. This laser was capable of delivering up to 4 J of energy in a single pulse, at a rate of one pulse per second, although pulses of only 0.5 J were used in our experiment. (This pulse energy, the time-integrated power of the yellow trace in Fig. 3, was confirmed with a Molectron JMAX 43 pyroelectric joulemeter.) We estimate that $\sim$0.3 J was absorbed in the maser crystal per pulse. With the frequency of the $TE_{01\delta}$ mode set near to the centre of the X$\leftrightarrow$Z spin transition, each pump pulse absorbed by the crystal produced a vigorous burst of maser oscillation (Fig. 3). The oscillation's magnitude as a function of the $TE_{01\delta}$ mode's frequency is shown in Fig. 4. The instantaneous output power of the maser burst (Fig. 3, blue trace), as monitored directly by a microwave receiver in the form of an HP8593E spectrum analyser set to fixed-frequency mode, was observed to exceed -10 dBm. Over 1,000 pump pulses, no change in the size or shape of the maser burst was noticeable provided that the resonator was kept in tune. By inserting neutral-density filters into the pump beam to extinguish the maser oscillation, the threshold pump energy exiting from the laser was determined to be 80 mJ over an effective pump pulse duration of 350 µs, which corresponds to a threshold pump power of about 230 W. Just above threshold, a considerably diminished maser burst, delayed in time and somewhat shorter in duration, could be observed.

We have constructed a simple model for how both the spin populations in the X and Z sublevels and the energization of (that is, photon number in) the maser mode change as functions of time in response to the applied optical pump pulse (Methods). The maser output power predicted by this model was fitted to the experimentally observed maser burst, yielding the green curve in Fig. 3, together with estimates of three spin-dynamics parameters. Our modelling suggests that near-zero-field operation is particularly advantageous for maser action (masing): the rate of spin–lattice relaxation is low and the transition linewidth is narrow[27].

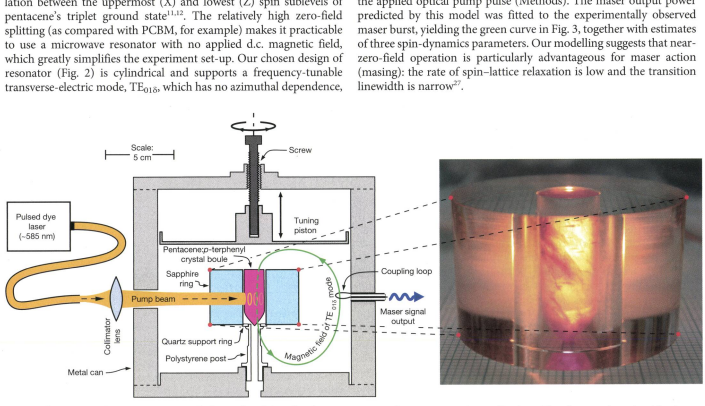


Figure 2 | Anatomy of the maser. A crystal of *p*-terphenyl doped with pentacene is located in the a.c. magnetic field of the $TE_{01\delta}$ mode of a microwave resonator and illuminated with a beam of yellow light from a pulsed dye laser. By maser action, the $TE_{01\delta}$ mode is energized, and from this mode a signal is extracted using a magnetic coupling loop. The photograph on the right shows the resonator's sapphire ring and the pentacene:*p*-terphenyl crystal slotted inside it; yellow back lighting reveals flaws within the crystal.

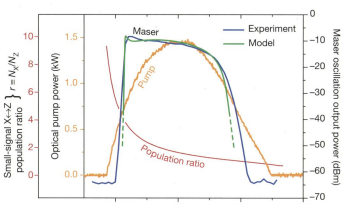


Figure 3 | Maser response in the time domain. The yellow trace (scale on near left) is the pump laser's instantaneous output power, which corresponds to an energy of ~0.5 J per pulse. The blue trace (scale on right) is the instantaneous power of the maser's oscillation burst as monitored with a microwave receiver at a resolution bandwidth of 10 MHz centred on 1.45 GHz. The green trace (same scale on right) is our fitted model's prediction of this instantaneous power. The red curve (scale on far left) is the predicted population ratio between the uppermost (X) and lowest (Z) maser levels in response to the optical pump in the low-signal regime.

Our experimental maser oscillator (Fig. 2) can be reconfigured as a pulsed-mode amplifier of weak electromagnetic signals by connecting its coupling loop to one port of a circulator[5]. A detailed, quantitative analysis of this amplifier's noise performance, taking into account deleterious sources of noise associated with losses within this circulator and the maser cavity itself, is provided in Methods. The red curve in Fig. 3 shows the population ratio $r \equiv N_X/N_Z$ between the uppermost (X) and lowest (Z) maser levels as predicted by our fitted model in response to the optical pump pulse shown (yellow trace). Whereas the blue and green curves correspond to saturation-limited maser oscillation, this red curve corresponds to the low-signal regime, where the energization of the maser mode is miniscule. Although decaying with time, r remains in excess of 2 for up to 150 μs from the onset of the pump pulse, implying that the intrinsic limit on the maser amplifier's residual noise temperature stays below 140 mK for the same amount of time. Whether or not r can be sustained sufficiently above 1 to allow continuous maser operation (given a suitable pump source and adequate cooling) would seem to be finely balanced with respect to the values of the relevant relaxation parameters.

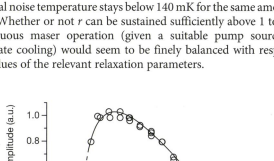


Figure 4 | Frequency response of maser action. The circles and associated guide line plot the measured plateau amplitude of the maser oscillation (vertical scale, obtained by down-converting the oscillation's output to a few tens of kilohertz and observing the resultant beat on a digital storage oscilloscope; a.u., arbitrary units) on illuminating the pentacene:*p*-terphenyl crystal (within its sapphire ring) with a ~0.5-J pulse of yellow light, for different frequencies of the microwave resonator's TE_{018} mode (horizontal scale).

When down-converted, sampled and Fourier-transformed, each burst of maser oscillation had a linewidth of ~6 kHz, consistent with that of a perfect sine wave modulated by the maser burst's envelope (Fig. 3, blue and green curves). Precise measurements of the maser's noise will be done if and when continuous operation is achieved.

With this device, we have demonstrated zero-field maser operation at room temperature with a solid-state gain medium. By improving the quality of the maser crystal, the effectiveness of the optical pumping and the Purcell factor of the resonator, we believe that the maser's threshold pump power can be substantially reduced, its efficiency (and, hence, output power as an oscillator) further increased and its residual noise temperature as an amplifier reduced. The application of a d.c. magnetic field bias should allow the maser's operating frequency to be tuned somewhat. The prospect of continuous maser operation, although not resolved here, remains a possibility and motivates attempts to identify different (yet perhaps structurally related) molecular systems whose properties are yet more advantageous for masing than those of pentacene:*p*-terphenyl.

Extremely low-noise amplifiers allow the detection of weak yet potentially important electromagnetic signals, which would otherwise go unnoticed. Low-phase-noise oscillators are essential components within frequency standards and clocks. Ultrasensitive parametric probes and detectors can, moreover, be derived from such oscillators. Because it can be operated at room temperature and in weak ambient magnetic fields (such as Earth's), the maser reported here thus opens new routes to the making of measurements at unprecedented levels of sensitivity and precision in areas such as space exploration, molecular biology and quantum computing. Furthermore, the general exploitation of intersystem crossing as an effective source of spin-polarized 'fuel' for the powering of more intricately orchestrated quantum molecular processes (beyond merely that of masing through a common microwave field) would seem to be a most fertile area of study.

METHODS SUMMARY

The maser crystal was grown using the vertical Bridgman–Stockbarger method in a double-walled glass vial[29]. Its constituent *p*-terphenyl was moderately zone-refined before growth; the substitutional concentration of its pentacene dopant was approximately 0.01%. The as-grown crystal boule was trimmed and then polished. It had extensive twinning (as revealed by spatially inhomogeneous birefringence) and other internal flaws (Fig. 2).

The microwave resonator was designed using both mode-matching and finite-element-method[28] software. The optical (*c*) axis of its monocrystalline sapphire ring lay parallel to the ring's geometric axis of rotation to within a tolerance of ± 1°. All of the ring's surfaces and the internal surfaces of the resonator's silver-plated metal can were polished. On inserting the pentacene:*p*-terphenyl boule into the ring, the X↔Z transition was initially located by directly monitoring on an oscilloscope the response of an L-band Robinson oscillator bridge (which incorporated the resonator within its microwave circuit) to pulses of below-threshold yellow pump light.

The spin populations in the triplet ground state's X and Z sublevels, together with the photon number in the 1.45-GHz TE_{018} microwave mode, were modelled by coupled linear rate equations capturing the processes of stimulated emission and absorption, spin–lattice relaxation, and intersystem-crossing-mediated population into, and decay out of, the triplet sublevels[30]; see Methods for further details, including estimates of parameters that enter these equations as coefficients.

The predicted noise performance of our maser is based on formulae derived in ref. 5, which have previously been applied without low-temperature approximations to cryogenic ruby masers[6]. Noise contributions associated with losses in passive microwave components (including the resonator), as well as the 'intrinsic' maser noise associated with the effective spin temperature across the X↔Z transition, are all included in our analysis.

Full Methods and any associated references are available in the online version of the paper.

Received 30 March; accepted 18 June 2012.

1. Schawlow, A. L. & Townes, C. H. Infrared and optical masers. *Phys. Rev.* **112,** 1940–1949 (1958).

2. Maiman, T. H. Stimulated optical radiation in ruby. *Nature* **187,** 493–494 (1960).
3. Kleppner, D., Goldenberg, H. M. & Ramsey, N. F. Properties of the hydrogen maser. *Appl. Opt.* **1,** 55–60 (1962).
4. Konoplev, I. V. *et al.* Experimental study of coaxial free-electron maser based on two-dimensional distributed feedback. *Phys. Rev. Lett.* **96,** 035002 (2006).
5. Siegman, A. E. *Microwave Solid-State Masers* (McGraw-Hill, 1964).
6. Clauss, R. C. & Shell, J. S. Ruby masers. In *Low-Noise Systems in the Deep Space Network* (ed. Reid, M. S.) (Deep Space Communication and Navigation Series, Jet Propulsion Laboratory, Caltech, 2008).
7. Dick, G. J. & Wang, R. T. Ultra-stable performance of the superconducting cavity maser. *IEEE Trans. Instrum. Meas.* **40,** 174–177 (1991).
8. Bourgeois, P.-Y. *et al.* Maser oscillation in a whispering-gallery-mode microwave resonator. *Appl. Phys. Lett.* **87,** 224104 (2005).
9. Mollier, J. C., Hardin, J. & Uebersfeld, J. Theoretical and experimental sensitivities of ESR spectrometers using maser techniques. *Rev. Sci. Instrum.* **44,** 1763–1771 (1973).
10. van der Waals, J. H. EPR of photo-excited triplet states: a personal account. *Appl. Magn. Reson.* **20,** 545–561 (2001).
11. Sloop, D. J., Yu, H.-L., Lin, T.-S. & Weissman, S. I. Electron spin echoes of a photoexcited triplet: pentacene in *p*-terphenyl crystals. *J. Chem. Phys.* **75,** 3746–3757 (1981).
12. Köhler, J., Brouwer, A. C. J., Groenen, E. J. J. & Schmidt, J. On the intersystem crossing of pentacene in *p*-terphenyl. *Chem. Phys. Lett.* **250,** 137–144 (1996).
13. Bogle, G. S. & Symmons, H. F. Zero-field masers. *Aust. J. Phys.* **12,** 1–20 (1959).
14. Whelehan, J. J. Low-noise amplifiers—then and now. *IEEE Trans. Microw. Theory Tech.* **50,** 806–813 (2002).
15. Wadefalk, N. *et al.* Cryogenic wide-band ultra-low-noise IF amplifiers operating at ultra-low DC power. *IEEE Trans. Microw. Theory Tech.* **51,** 1705–1711 (2003).
16. Smith, B. A. *et al.* A new look at the Saturn system: the Voyager 2 images. *Science* **215,** 504–537 (1982).
17. Asztalos, S. J. *et al.* SQUID-based microwave cavity search for dark-matter axions. *Phys. Rev. Lett.* **104,** 041301 (2010).
18. Korolev, A. M., Shnyrkov, V. I. & Shulga, V. M. Ultra-high frequency ultra-low dc power consumption HEMT amplifier for quantum measurements in millikelvin temperature range. *Rev. Sci. Instrum.* **82,** 016101 (2011).
19. Blank, A., Kastner, R. & Levanon, H. Exploring new active materials for low-noise room-temperature microwave amplifiers and other devices. *IEEE Trans. Microw. Theory Tech.* **46,** 2137–2144 (1998).
20. Hsu, H. & Tittel, F. K. Optical pumping of microwave masers. *Proc. IEEE* **51,** 185–189 (1963).
21. Sabisky, E. S. & Anderson, C. H. A solid-state cw optically pumped microwave maser. *Appl. Phys. Lett.* **8,** 298–300 (1966).
22. Blank, A. & Levanon, H. Toward maser action at room temperature by triplet-radical interaction and its application to microwave technology. *RIKEN Rev.* **44,** 128–130 (2002).
23. Blank, A. & Levanon, H. Applications of photoinduced electron spin polarization at room temperature to microwave technology. *Appl. Phys. Lett.* **79,** 1694–1696 (2001).
24. Antipov, S. *et al.* in *Proc. 23rd Particle Accelerator Conf.* 4535–4537, https://accelconf.web.cern.ch/accelconf/PAC2009/papers/fr5rfp001.pdf (Joint Accelerator Conference Website, 2009).
25. Kagawa, A., Murokawa, Y., Takeda, K. & Kitagawa, M. Optimization of ^{3}He spin density for dynamic nuclear polarization using photo-excited triplet electron spins. *J. Magn. Reson.* **197,** 9–13 (2009).
26. Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69,** 681–686 (1946).
27. Lang, J., Sloop, D. J. & Lin, T.-S. Orientational anisotropic studies by field rotation technique: near zero-field pulsed EPR experiments of pentacene in *p*-terphenyl. *J. Magn. Reson.* **176,** 249–256 (2005).
28. Selvakumar, S. *et al.* Growth and studies on SSVBT grown *p*-terphenyl single crystals. *J. Cryst. Growth* **275,** e265–e271 (2005).
29. Oxborrow, M. Traceable 2-D finite-element simulation of the whispering-gallery modes of axisymmetric electromagnetic resonators. *IEEE Trans. Microw. Theory Tech.* **55,** 1209–1218 (2007).
30. Takeda, K., Takegoshi, K. & Terao, T. Zero-field electron spin resonance and theoretical studies of light penetration into single crystals and polycrystalline material doped with molecules photoexcitable to the triplet state via intersystem crossing. *J. Chem. Phys.* **117,** 4940–4946 (2002).

Acknowledgements M.O. thanks J. Sherwood, R. Ferguson and C. Langham for advice on the growth of molecular crystals, for help in cutting and polishing the crystalline *p*-terphenyl boule and for designing the metal components of the microwave resonator, respectively. M.O. and J.D.B. thank C. Kay for discussions and guidance on pulsed electron paramagnetic resonance spectroscopy. This work was supported by the NMS Pathfinder Metrology Programme and by the Engineering and Physical Sciences Research Council.

Author Contributions J.D.B. performed initial pulsed electron paramagnetic resonance experiments with M.O., and analysed the resultant data. Both J.D.B. and M.O. carried out independent electromagnetic modelling of the TE_{015}-mode microwave resonator using different software. M.O. grew the pentacene:*p*-terphenyl crystal, designed and constructed the microwave circuitry and optical systems, performed the final experiments and analysed the resultant data. N.M.A. initiated the original work on high-*Q* cavities. N.M.A. and M.O. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.O. (mo@npl.co.uk) or N.M.A. (n.alford@imperial.ac.uk).

LETTER

doi:10.1038/nature11341

Electronic read-out of a single nuclear spin using a molecular spin transistor

Romain Vincent[1], Svetlana Klyatskaya[2], Mario Ruben[2,3], Wolfgang Wernsdorfer[1] & Franck Balestro[1]

Quantum control of individual spins in condensed-matter devices is an emerging field with a wide range of applications, from nanospintronics[1,2] to quantum computing[3]. The electron, possessing spin and orbital degrees of freedom, is conventionally used as the carrier of quantum information in proposed devices[4,5,6,7,8,9]. However, electrons couple strongly to the environment, and so have very short relaxation and coherence times. It is therefore extremely difficult to achieve quantum coherence and stable entanglement of electron spins. Alternative concepts propose nuclear spins as the building blocks for quantum computing[10], because such spins are extremely well isolated from the environment and less prone to decoherence. However, weak coupling comes at a price: it remains challenging to address and manipulate individual nuclear spins[11,12,13,14]. Here we show that the nuclear spin of an individual metal atom embedded in a single-molecule magnet can be read out electronically. The observed long lifetimes (tens of seconds) and relaxation characteristics of nuclear spin at the single-atom scale open the way to a completely new world of devices in which quantum logic may be implemented.

Single-shot read-out of a single nuclear spin has remained elusive until the past few years, when it was achieved on a single nitrogen-vacancy defect centre in diamond[15]. The use of single-molecule magnets (SMMs) has also been proposed as a structural basis for nanospintronics[16] and quantum information[9]. The argument in favour of molecular building blocks for quantum computing is based on the possibility of synthesizing billions of perfectly identical SMMs, which can be integrated in devices by bottom-up approaches and for which the intrinsic magnetic properties and the environment can be chemically engineered. Moreover, experiments on an assembly of SMMs have already shown that: SMMs are characterized by a large spin ground state combined with uniaxial magnetic anisotropy leading to two stable spin orientations; quantum phenomena such as quantum tunnelling of the magnetization (QTM)[17,18] and quantum phase interference[19] can be observed; and spin-state oscillations can be made coherent[20,21] for a time on the order of 1 μs. To achieve a longer coherence time, experiments must be done on the single-molecule level, avoiding interactions between SMMs. However, it is still difficult to detect single spin states in an SMM. Towards this goal, transport measurements have been made using scanning tunnelling microscopy (STM)[22,23], and SMMs have been studied in a transistor-like configuration using electromigrated junctions[24,25] or carbon-nanotube transistors[26]. So far, only signatures accounting for the electronic magnetic moment have been addressed, and there has been no quantitative comparison with the expected theoretical magnetic behaviour of an individual SMM.

Here, transport measurements taken through a single bis(phthalocyaninato)terbium(III) SMM ($TbPc_2$, Fig. 1) are studied in a three-terminal geometry obtained by electromigration[27]. $TbPc_2$ is a lanthanide SMM, in which the magnetic moment is carried by a single Tb^{3+} metal ion sandwiched between two organic phthalocyanine (Pc) ligands. We used this particular SMM for the several reasons. It has been reported that $TbPc_2$ SMMs conserve both their structural integrity and their magnetic properties even when sublimated at 820 K on a copper surface[28]. The redox state of the Tb^{3+} ion is very stable; that is, it is very unlikely that a current can flow through the Tb^{3+} ion (see Supplementary Information). The two Pc ligands have a conjugated π system, which can easily conduct electrons. Owing to strong spin–orbit coupling and the Pc ligand fields, $TbPc_2$ has a magnetic ground state of $J = 6$ and a strong uniaxial magnetic anisotropy (Ising-like, Fig. 1a). The ground-state doublet $J_z = \pm 6$ is well isolated

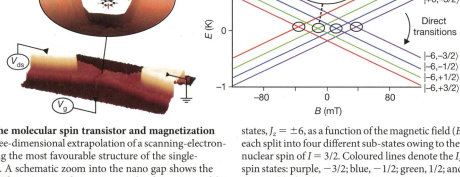


Figure 1 | Geometry of the molecular spin transistor and magnetization reversal processes. a, Three-dimensional extrapolation of a scanning-electron-microscope image showing the most favourable structure of the single-molecule-based transistor. A schematic zoom into the nano gap shows the molecular structure of the $TbPc_2$ SMM and its easy axis. The charge state of the ligand read-out dot can be controlled by the gate voltage, V_g, and the voltage difference between the electrodes is controlled through the drain-source voltage, V_{ds}. **b**, Zeeman diagram presenting the energy, E, of the two ground states, $J_z = \pm 6$, as a function of the magnetic field (B). The two ground states are each split into four different sub-states owing to the hyperfine coupling with the nuclear spin of $I = 3/2$. Coloured lines denote the I_z components of the nuclear spin states: purple, −3/2; blue, −1/2; green, 1/2; and red, 3/2. Two processes are responsible for the magnetization reversal. In small magnetic fields, QTM can occur at the avoided energy-level crossings with the same I_z but opposite J_z, indicated by the black circles. In higher fields, a direct relaxation process can lead to the reversal of J.

[1]Institut Néel, CNRS et Université Joseph Fourier, BP 166, F-38042 Grenoble Cedex 9, France. [2]Institute of Nanotechnology, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany. [3]Institut de Physique et Chimie des Matériaux de Strasbourg, CNRS-Université de Strasbourg, 67034 Strasbourg, France.

from the excited states by an energy gap of a few hundred kelvin[29]. The excited states can thus be neglected for experiments carried out at low temperatures (less than 5 K) and small magnetic fields (less than 10 T). The ligand fields also induce transverse magnetic anisotropy, which is responsible for the coupling of the two states of the doublet $J_z = \pm 6$, so that energy-level crossings are avoided[29] (see the black circles in the Zeeman diagram of Fig. 1b). Finally, because of strong hyperfine coupling with the nuclear spin $I = 3/2$ of the Tb^{3+} ion, each ground state $J_z = \pm 6$ is split into four states (Fig. 1b).

Within the low-temperature regime, the magnetization reversal of a $TbPc_2$ SMM takes place through two different processes. In small magnetic fields, QTM at the four avoided level crossings (see above) is the main mechanism[29], whereas in higher magnetic fields, relaxation processes can be direct, involving non-coherent tunnelling events combined with phonon emission. QTM of a $TbPc_2$ SMM has been observed in a supramolecular carbon-nanotube SMM spin-valve geometry[26], but the signature of the hyperfine coupling could not be accessed, owing to the background noise induced by the surrounding $TbPc_2$ SMMs. Here, we show that by detecting the reversal of the electronic magnetic moment of the $TbPc_2$ SMM, we can produce an electronic read-out of the four nuclear spin states of a single Tb^{3+} ion with high efficiency and in perfect quantitative agreement with the theoretical predictions[29]. Furthermore, we estimate the lifetimes and the temperature of a single nuclear spin.

Transport measurements through a single $TbPc_2$ SMM have been taken previously by scanning tunnelling spectroscopy (STS) experiments[22], for which electronic transport occurred through the Pc ligands and exhibited Coulomb blockade and Kondo effects depending on its charge state (spin states $S = 0$ or 1/2). However, no signature of the magnetic moment carried by the Tb^{3+} ion was observed in this experiment. To detect the reversal of the magnetic moment, and to produce the electronic read-out of the Tb^{3+} nuclear spin states, we inserted the $TbPc_2$ SMM directly into a gold junction obtained by the electromigration technique (see Fig. 1a and Methods). The differential conductance, dI/dV, as the function of the drain-source voltage, V_{ds}, and the gate voltage, V_g, is presented in Fig. 2a. It exhibits a single charge-degeneracy point with a weak spin $S = 1/2$ Kondo effect. A detailed study of the Kondo peak as a function of the applied magnetic field is presented in the Supplementary Information. A ferromagnetic exchange interaction of about 0.35 T was measured between the spin 1/2 of the quantum dot and the magnetic moment carried by the Tb^{3+} ion. Alternative coupling mechanisms (dipolar, magneto-Coulomb, mechanical, flux) are discussed in the Supplementary Information, but the relatively high value of the exchange interaction means that it should be the major contribution to the coupling mechanism, and that the quantum dot is spatially located close to the Tb^{3+} ion. This is indirect proof that the electronic transport occurs through the aromatic Pc ligands, and that the most favourable geometric configuration is the one depicted in Fig. 1a. To summarize: the Pc ligands form a molecular quantum dot and the anisotropic magnetic moment of the Tb^{3+} ion is coupled to the electron path only indirectly, mainly through a ferromagnetic exchange interaction; the presence of a gate allows fine-tuning from the Coulomb blockade to the Kondo regime of the molecular quantum dot; the magnetic properties of the Tb^{3+} ion are then independent of the charge state of the Pc quantum dot (see Supplementary Information); and owing to the exchange coupling, we can use the Pc ligands as a read-out quantum dot to detect the reversal of the electronic magnetic moment carried by the Tb^{3+} ion spin dot.

To produce the electronic read-out of the single nuclear spin carried by the spin dot, we chose experimental conditions close to the charge-degeneracy point ($V_g = -0.9$ V and $V_{ds} = 0$ V in Fig. 2a, see Supplementary Information). When sweeping the magnetic field from negative to positive values, we observed a single abrupt jump of the differential conductance, which reversed when we swept the field in the opposite direction (Fig. 2b). Our detailed studies (see below) showed that these jumps can be accounted for by the reversals of the Tb^{3+} magnetic moment, which slightly influence the chemical potential of the read-out quantum dot through the magnetic interactions

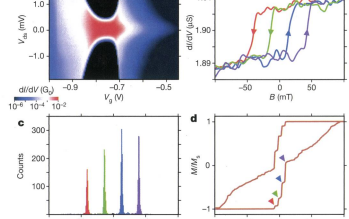


Figure 2 | Conductance characteristics and electronic read-out procedure. **a**, Stability diagram of the Pc read-out quantum dot exhibiting the differential conductance, dI/dV, in units of the quantum of conductance, G_0, as a function of gate voltage, V_g, and bias voltage, V_{ds}, at 0.1 K. **b**, dI/dV measurements for a given working point ($V_g = -0.9$ V; $V_{ds} = 0$ V) as function of the magnetic field, B. The arrows indicate the field-sweep direction. Abrupt jumps in the differential conductance, attributed to the switching of the Tb^{3+} magnetic moment, are visible for all traces of B, showing a clear hysteresis in the dI/dV characteristics. **c**, Histogram of switching field obtained for 11,000 field sweeps showing four preferential field values that are assigned to QTM events. **d**, Normalized hysteresis loop of a single $TbPc_2$ SMM obtained by integration of 1,000 field sweeps and performed for trace and retrace on a larger magnetic-field range than in **c**. The four arrows on the trace curve show the four preferential field values associated to QTM (red, −40 mT; green, −14 mT; blue, 14 mT; purple, 40 mT).

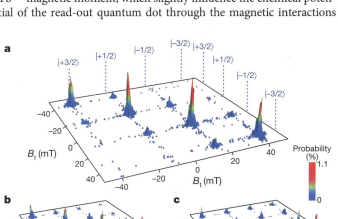


Figure 3 | Transition matrix of the QTM events as a function of the waiting time. The switching fields of the Tb^{3+} magnetic moment of subsequent field sweeps are plotted in two-dimensional histograms for three waiting times, t_w: **a**, $t_w = 0$ s; **b**, $t_w = 20$ s; and **c**, $t_w = 50$ s. The two axes correspond to the trace and retrace field sweeps, B_l and B_r, respectively. Two successive measurements with the same nuclear spin states are situated on the diagonal of the matrix, whereas the off-diagonal positions correspond to nuclear spin-state changes of $\Delta m_I = \pm 1, \pm 2$ and ± 3. The predominance of diagonal terms up to $t_w = 20$ s indicates the long level lifetime of the nuclear spin states. For $t_w = 50$ s, the diagonal terms vanish owing to nuclear spin-flip processes. Furthermore, the high amplitude of the bottom-right ($B_l = B_r = 40$ mT) matrix element accounts for the relaxation of the nuclear spin towards a thermal equilibrium.

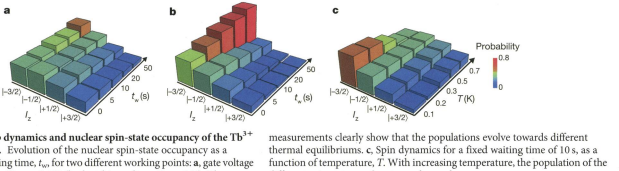


Figure 4 | Spin-flip dynamics and nuclear spin-state occupancy of the Tb^{3+} nuclear spin states. Evolution of the nuclear spin-state occupancy as a function of the waiting time, t_w, for two different working points: **a**, gate voltage $V_g = -0.9$ V; and **b**, $V_g = -0.1$ V (both at bias voltage $V_{ds} = 0$ V). The measurements clearly show that the populations evolve towards different thermal equilibriums. **c**, Spin dynamics for a fixed waiting time of 10 s, as a function of temperature, T. With increasing temperature, the population of the different spin states evolves towards equal occupancy.

(see Supplementary Information). We took subsequent measurements of the reversal as a function of the magnetic-field angle in a dilution refrigerator equipped with a home-made three-dimensional vector magnet. We were thus able to align the field-sweep direction accurately along the easy axis of the $TbPc_2$ SMM (see Supplementary Information).

We obtained the reversal statistics of the magnetization by sweeping the magnetic field back and forth at 0.05 T s^{-1} and measuring histograms of the switching fields. We present a typical example in Fig. 2c, showing four preferential magnetic-field values. These are in perfect quantitative agreement with theoretical predictions[29] of QTM of a Tb^{3+} magnetic moment at the four avoided energy-level crossings with the nuclear spin states $|3/2\rangle$, $|1/2\rangle$, $|-1/2\rangle$ and $|-3/2\rangle$ (Fig. 1b). Moreover, the histograms of the four switching fields do not overlap; this shows the high efficiency of this electronic read-out procedure. The field integration of the normalized switching histograms yields the magnetic hysteresis loop (Fig. 2d), which is in excellent accordance with micro-SQUID measurements of assemblies of $TbPc_2$ SMMs (see Supplementary Information and ref. 29). In higher magnetic fields, the reversal of the magnetization occurs stochastically, as predicted for a direct relaxation process involving a non-coherent tunnelling event combined with a phonon emission.

The lifetimes of the four nuclear spin states could be measured by studying the correlations between subsequent measurements as a function of the waiting time, t_w, between field sweeps. Figure 3 presents two-dimensional histograms (transition matrices) obtained from 22,000 field sweeps. The two axes correspond to the trace and retrace field sweeps, B_t and B_r. Two subsequent measurements with the same nuclear spin states are situated on the diagonal of the matrix; the off-diagonal positions correspond to nuclear spin state changes of $\Delta m_I = \pm 1, \pm 2$, and ± 3. For zero waiting time ($t_w = 0$ s, Fig. 3a), the diagonal positions are predominant, highlighting the robustness of the nuclear spin states; that is, indicating long level lifetimes for the individual Tb^{3+} nuclear spin states. The diagonal positions persist even for a waiting time of $t_w = 20$ s (Fig. 3b). However, for $t_w = 50$ s, the off-diagonal positions start to be populated, which suggests the occurrence of nuclear spin–flip processes during the waiting time (Fig. 3c). From this series of measurements, we conclude that the level lifetime of the nuclear spin states is on the order of tens of seconds, confirming that the invasiveness of the measurement procedure is low (see Supplementary Information).

More detailed insight into the spin–flip dynamics of an individual nuclear spin can be gained by measuring the population of nuclear spin states as a function of waiting time and temperature (Fig. 4). To this end, we determined the nuclear spin-state occupancy for two different working points corresponding to two different charge states of the read-out quantum dot ($V_g = -0.9$ V in Fig. 4a and $V_g = -0.1$ V in Fig. 4b, both at $V_{ds} = 0$ V). It is clear that the state occupancy, relaxing towards a thermal equilibrium, depends heavily on the transport characteristics (the current flowing through the Pc quantum dot and/or the electrostatic environment modulated by the gate voltage). Indeed, the electron tunnelling through the Pc-ligands read-out dot gives rise to small fluctuations of the local electric field, which could modify nuclear spin–flip processes through the quadrupole interaction[30]. A detailed study addressing this problem is in progress, and beyond the scope of the work presented here. We also determined the population of nuclear spin states as a function of the temperature (Fig. 4c). The strong temperature dependence of this occupancy demonstrates that a single nuclear spin can be thermalized down to at least 0.2 K, which is close to the electronic temperature of our dilution refrigerator (0.08 K).

In conclusion, we have performed electronic transport studies on a $TbPc_2$ SMM in a transistor-like set-up, taking measurements of a single nuclear spin. Our experimental read-out procedure relies on the highly efficient detection of the QTM of the electronic magnetic moment at particular values of the magnetic field corresponding to the four avoided energy-level crossings; this is in perfect quantitative agreement with theoretical predictions. Using this procedure, the dynamics of the four different nuclear spin states of a single Tb^{3+} ion could be electronically determined. The observation of energy-level lifetimes on the order of tens of seconds opens the way to coherent manipulations of a single nuclear spin. Moreover, because of the different energy-level spacing that originates from the nuclear quadrupole interaction term (see Supplementary Information), this particular device is presumed to be an excellent candidate to perform the Grover algorithm. Indeed it has been proposed that the coherent manipulation of a nuclear spin $I = 3/2$ can be performed through a multiphoton process using this algorithm[31]. It was predicted that the greater the quadrupolar constant, the better the implementation of the algorithm would be. Note that the value presented in ref. 31 was three orders of magnitude smaller than the value presented here ($P = 0.013$ cm^{-1}). Thus, the demonstrated possibility of addressing and detecting single nuclear spin states using the QTM of SMMs, in combination with the observation of long energy-level lifetimes, opens up a bright new world of nanospintronics with integrated memory, logics and possibly quantum logics.

METHODS SUMMARY

The single-molecule transistor was prepared by blow-drying a dilute toluene solution of the $TbPc_2$ molecule onto a gold nanowire on an Au/HfO_2 gate fabricated through atomic-layer deposition. Before the solution was blow-dried, the electrodes were cleaned with acetone, ethanol, isopropanol solution and oxygen plasma. The connected sample was enclosed in a high-frequency, low-temperature filter, consisting of a thermocoax microwave filter and π filters, anchored to the mixing chamber of a dilution refrigerator with a base temperature of about 0.05 K. The molecule-coated nanowire was then broken by electromigration, using a voltage ramp at 4 K. The electromigration technique used real-time electronics to increase the yield of coupling a single molecule to the electrodes.

Transport measurements were taken using a lock-in amplifier in a dilution refrigerator with an electronic temperature of about 0.08 K. It was equipped with a home-made three-dimensional vector magnet, allowing us to sweep the magnetic field in three dimensions at rates up to 0.2 T s^{-1}.

Received 9 April; accepted 19 June 2012.

1. Awschalom, D. D., Loss, D. & Samarth, N. *Semiconductor Spintronics and Quantum Computation* (Springer, 2002).

2. Zutic, I., Fabian, J. & Das Sarma, S. Spintronics: fundamentals and applications. *Rev. Mod. Phys.* **76,** 323–410 (2004).
3. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
4. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57,** 120–126 (1998).
5. Imamoglu, A. *et al.* Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83,** 4204–4207 (1999).
6. Eble, B. *et al.* Hole–nuclear spin interaction in quantum dots. *Phys. Rev. Lett.* **102,** 1–4 (2009).
7. Jelezko, F., Gaebel, T., Popa, I., Gruber, A. & Wrachtrup, J. Observation of coherent oscillations in a single electron spin. *Phys. Rev. Lett.* **92,** 1–4 (2004).
8. Vrijen, R. *et al.* Electron-spin-resonance transistors for quantum computing in silicon-germanium heterostructures. *Phys. Rev. A.* **62,** 1–10 (2000).
9. Leuenberger, M. N. & Loss, D. Quantum computing in molecular magnets. *Nature* **410,** 789–793 (2001).
10. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393,** 133–137 (1998).
11. Chuang, I. L., Gershenfeld, N. & Kubinec, M. Experimental implementation of fast quantum searching. *Nature* **393,** 143–146 (1998).
12. Chuang, I. L., Vandersypen, L. M. K., Zhou, X., Leung, D. W. & Lloyd, S. Experimental realization of a quantum algorithm. *Phys. Rev. Lett.* **80,** 3408–3411 (1998).
13. DiVincenzo, D. P. Two-bit gates are universal for quantum computation. *Phys. Rev. A.* **51,** 1015–1022 (1995).
14. Berman, G. P., Doolen, G. D., Hammel, P. C. & Tsifrinovich, V. I. Solid-state nuclear-spin quantum computer based on magnetic resonance force microscopy. *Phys. Rev. B* **61,** 14694–14699 (2000).
15. Neumann, P. *et al.* Single-shot readout of a single nuclear spin. *Science* **329,** 542–544 (2010).
16. Bogani, L. & Wernsdorfer, W. Molecular spintronics using single-molecule magnets. *Nature Mater.* **7,** 179–186 (2008).
17. Friedman, J. R., Sarachik, M. P., Tejada, J. & Ziolo, R. Macroscopic measurement of resonant magnetization tunnelling in high-spin molecules. *Phys. Rev. Lett.* **76,** 3830–3833 (1996).
18. Thomas, L. *et al.* Macroscopic quantum tunnelling of magnetization in a single crystal of nanomagnets. *Nature* **383,** 145–147 (1996).
19. Wernsdorfer, W. & Sessoli, R. Quantum phase interference and parity effects in magnetic molecular clusters. *Science* **284,** 133–135 (1999).
20. Ardavan, A. *et al.* Will spin-relaxation times in molecular magnets permit quantum information processing? *Phys. Rev. Lett.* **98,** 1–4 (2007).
21. Bertaina, S. *et al.* Quantum oscillations in a molecular magnet. *Nature* **453,** 203–206 (2008).
22. Komeda, T. *et al.* Observation and electric current control of a local spin in a single-molecule magnet. *Nature Commun.* **2,** 217 (2011).
23. Otte, A. F. *et al.* The role of magnetic anisotropy in the Kondo effect. *Nature Phys.* **4,** 847–850 (2008).
24. Heersche, H. B. *et al.* Electron transport through single Mn_{12} molecular magnets. *Phys. Rev. Lett.* **96,** 206801 (2006).
25. Zyazin, A. S. *et al.* Electric field controlled magnetic anisotropy in a single molecule. *Nano Lett.* **10,** 3307–3311 (2010).
26. Urdampilleta, M., Cleuziou, J.-P., Klyatskaya, S., Ruben, M. & Wernsdorfer, W. Supramolecular spin valves. *Nature Mater.* **10,** 502–506 (2011).
27. Park, H., Lim, A. K. L., Alivisatos, A. P., Park, J. & McEuen, P. L. Fabrication of metallic electrodes with nanometer separation by electromigration. *Appl. Phys. Lett.* **75,** 301–303 (1999).
28. Stepanow, S. *et al.* Spin and orbital magnetic moment anisotropies of monodispersed bis(phthalocyaninato)terbium on a copper surface. Supp. Inf. *J. Am. Chem. Soc.* **132,** 11900–11901 (2010).
29. Ishikawa, N., Sugita, M. & Wernsdorfer, W. Quantum tunneling of magnetization in lanthanide single-molecule magnets: bis(phthalocyaninato)terbium and bis(phthalocyaninato)dysprosium anions. *Angew. Chem. Int. Ed.* **44,** 2931–2935 (2005).
30. Abragam, A. *The Principles of Nuclear Magnetism* (Oxford Univ. Press, 1994).
31. Leuenberger, M. N. & Loss, D. The Grover algorithm with large nuclear spins in semiconductors. *Phys. Rev. B* **68,** 165317 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Eyraud, D. Lepoittevin, L. Cagnon, R. Haettel, C. Hoarau, V. Reita and P. Rodière for technical contributions and discussions. We thank T. Fournier, T. Crozes, B. Fernandez, S. Dufresnes and G. Julie for lithography development; E. Bonet, C. Thirion and R. Piquerel for help with software development and M. Urdampilleta, S. Thiele, N. Roch, A. Variet and A. Ralko for discussions. Samples were fabricated in the Nanofab facility of the Néel Institute. This work is partially supported by the French National Research Agency National Programme in Nanosciences and Nanotechnologies (ANR-PNANO) project MolNanoSpin, number ANR-08-NANO-002; European Research Council Advanced Grant MolNanoSpin, number 226558; ICT-2007.8.0 Future Emerging Technologies Open, Quantum Information Processing Specific Targeted Research Project number 211284 MolSpinQIP; the German Research Foundation programme TRR 88 '3Met'; Cible 2009; and the Nanosciences Foundation of Grenoble.

Author Contributions R.V., W.W. and F.B. designed, conducted and analysed the experiments; S.K. and M.R. designed, synthesized and characterized the molecule; R.V., M.R., W.W., and F.B. co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.B. (franck.balestro@grenoble.cnrs.fr) or M.R. (mario.ruben@kit.edu).

LETTER

doi:10.1038/nature11338

Doubling of marine dinitrogen-fixation rates based on direct measurements

Tobias Großkopf[1]*, Wiebke Mohr[1]*†, Tina Baustian[1], Harald Schunck[1], Diana Gill[1], Marcel M. M. Kuypers[2], Gaute Lavik[2], Ruth A. Schmitz[3], Douglas W. R. Wallace[4] & Julie LaRoche[1]†

Biological dinitrogen fixation provides the largest input of nitrogen to the oceans, therefore exerting important control on the ocean's nitrogen inventory and primary productivity[1–3]. Nitrogen-isotope data from ocean sediments suggest that the marine-nitrogen inventory has been balanced for the past 3,000 years (ref. 4). Producing a balanced marine-nitrogen budget based on direct measurements has proved difficult, however, with nitrogen loss exceeding the gain from dinitrogen fixation by approximately 200 Tg N yr^{-1} (refs 5, 6). Here we present data from the Atlantic Ocean and show that the most widely used method of measuring oceanic N_2-fixation rates[7] underestimates the contribution of N_2-fixing microorganisms (diazotrophs) relative to a newly developed method[8]. Using molecular techniques to quantify the abundance of specific clades of diazotrophs in parallel with rates of $^{15}N_2$ incorporation into particulate organic matter, we suggest that the difference between N_2-fixation rates measured with the established method[7] and those measured with the new method[8] can be related to the composition of the diazotrophic community. Our data show that in areas dominated by *Trichodesmium*, the established method underestimates N_2-fixation rates by an average of 62%. We also find that the newly developed method yields N_2-fixation rates more than six times higher than those from the established method when unicellular, symbiotic cyanobacteria and γ-proteobacteria dominate the diazotrophic community. On the basis of average areal rates measured over the Atlantic Ocean, we calculated basin-wide N_2-fixation rates of 14 ± 1 Tg N yr^{-1} and 24 ± 1 Tg N yr^{-1} for the established and new methods, respectively. If our findings can be extrapolated to other ocean basins, this suggests that the global marine N_2-fixation rate derived from direct measurements may increase from 103 ± 8 Tg N yr^{-1} to 177 ± 8 Tg N yr^{-1}, and that the contribution of N_2 fixers other than *Trichodesmium* is much more significant than was previously thought.

The oceans support roughly half of Earth's biological carbon fixation[9]. Carbon itself is rarely considered to be limiting in the ocean; instead, elements such as nitrogen, phosphorus and iron control primary productivity[2]. Of these elements, nitrogen is special because it is always present in large amounts as dissolved N_2, but this form is available only to the diazotrophs, a restricted group of prokaryotes[1]. Biological N_2 fixation is the largest source of fixed-nitrogen input into the ocean and it has been proposed that this controls marine primary productivity on geological timescales[10]. Whereas phytoplankton seem to be limited in growth over large areas by the availability of fixed N (for example, nitrate, nitrite or ammonium), the productivity of diazotrophs is controlled by other environmental factors, such as the availability of Fe and P (ref. 11). Thus, increased desertification owing to changes in land use may promote marine N_2 fixation through increased aeolian input of iron[12], whereas the increased use of fertilizers and riverine run-off or increased atmospheric N deposition may have the opposite effect[13,14]. An understanding of marine N_2 fixation and its response to anthropogenic forcing is crucial for assessing the future of oceanic primary productivity. However, recent attempts to produce a balanced N budget have usually left a major gap on the input side[5,6]. This creates a paradox, because stable isotopes recorded in ocean sediments suggest that the marine N inventory has been balanced for the past 3,000 years (ref. 4). The paradox can be reconciled in three ways[6]. One hypothesis is that current, transient N-loss rates are exceeding N_2 fixation, but that over timescales of several thousand years, variable losses and gains act to maintain a steady-state fixed-N inventory[15,16]. Two alternate hypotheses invoke the possibilities that either the N-loss or the N-gain terms are in error. Several lines of evidence point out that even the most conservative estimates of oceanic N losses are higher than current estimates of N input based on field measurements of N_2 fixation. This suggests that contemporary N_2-fixation rates may have been underestimated grossly[5,6].

In the past decade, molecular techniques have led to the discovery of a variety of previously unrecognized diazotrophs that express their nitrogenase genes and fix N_2, thereby showing that there are gaps in our knowledge of marine N_2 fixation[17–21]. The discovery of N_2 fixation in mesopelagic waters in the Pacific[22,23] and the possibility of significant involvement of heterotrophic bacteria in global marine N_2 fixation[24,25] reinforce the idea that we are missing a large fraction of global marine N input by N_2 fixation. Biogeochemical modelling and geochemical-tracer techniques have been used to estimate regional and global magnitudes of marine N_2 fixation, but these indirect approaches rely on assumptions that require verification by direct biological rate measurements of N_2 fixation.

A recent laboratory study showed that the established method for measuring oceanic N_2 fixation (accounting for around 90% of the published rates[26]) leads to an underestimate, owing to slow and incomplete equilibration between the water sample and the $^{15}N_2$ tracer added as a gas bubble[8]. If this finding holds for measurements of environmental N_2-fixation rates, an upward revision of the largest source of combined N into the ocean—that is, N_2-fixation—would be inevitable. We investigated the magnitude of the underestimation of N_2-fixation rates in open-ocean settings. In autumn 2009, on two research cruises to the Atlantic Ocean between 25° N and 45° S, we compared the established $^{15}N_2$ tracer method for measuring N_2 fixation[7] (hereafter called the bubble-addition method) with a recently developed method[8] in which the $^{15}N_2$ tracer is added as a dissolved gas (the dissolution method). Our sampling covered a wide variety of oceanic conditions with sea surface temperatures ranging from 10 °C to more than 28 °C (Fig. 1). At each station, we conducted parallel incubations with the dissolution and the bubble-addition methods, using dual labelling with $NaH^{13}CO_3$ and $^{15}N_2$ gas in the same incubation bottles to measure

[1]Helmholtz Centre for Ocean Research Kiel (GEOMAR), Düstembrooker Weg 20, 24105 Kiel, Germany. [2]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany. [3]Institute for General Microbiology, Christian-Albrechts University Kiel, Am Botanischen Garten 1–9, 24118 Kiel, Germany. [4]Oceanography Department, Dalhousie University, 1355 Oxford Street, PO Box 15000, Halifax, Nova Scotia, B3H 4R2, Canada. †Present addresses: Harvard University, Department of Earth and Planetary Sciences, 20 Oxford Street, Cambridge, Massachusetts 02138, USA (W.M.); Department of Biology, Dalhousie University, 1355 Oxford Street, PO Box 1500, Halifax, Nova Scotia B3H 4R2, Canada (J.L.R.).
*These authors contributed equally to this work.

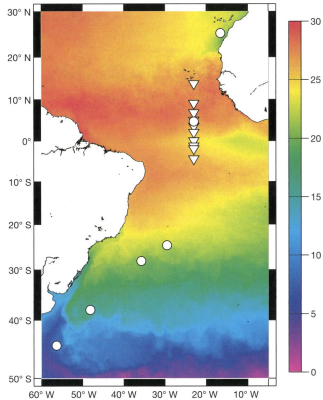


Figure 1 | Sampling sites and sea surface temperature. Map of average night-time sea surface temperature (SST) in °C in autumn 2009 in the Atlantic Ocean (http://oceancolor.gsfc.nasa.gov/), overlaid with stations sampled on the RV *Meteor* cruise 80–81 (triangles) and the RV *Polarstern* cruise ANT-XXVI/1 (circles). Triangles also correspond to 11 vertical profile stations used in Fig. 3. The equatorial station was sampled twice during the course of the cruise, so two triangles are superimposed at this location.

dissolved inorganic carbon (DIC) assimilation and N_2 fixation simultaneously with both methods (see Supplementary Information for experimental details).

The ^{13}C-labelled DIC was added to check for systematic differences in biological activity attributable to the differences in methods, and the results showed that carbon-fixation rates were in good agreement between the two experimental methods (Fig. 2b). By contrast, there was a large difference and poor overall correlation between the N_2-fixation rates calculated using the two methods (Fig. 2a). Average depth-integrated rates of N_2 fixation over the whole Atlantic Ocean (25° N to 45° S) differed by a factor of 1.7 (91 ± 4 and 54 ± 2 μmol N m^{-2} d^{-1} for the dissolution and bubble-addition methods, respectively; mean and propagated s.e.m., $n = 17$; Wilcoxon matched pairs test $P < 0.01$). Furthermore, a geographical pattern emerged among the differences in the rates measured with the two methods. In the northern part of the 23° W section (15° N to 5° S), the depth-integrated N_2-fixation rates derived from four vertical profiles were, on average, 62% higher with the dissolution method than with the bubble-addition method (194 ± 12 and 120 ± 5 μmol N m^{-2} d^{-1}, respectively; means and propagated s.e.m., $n = 4$). In the equatorial Atlantic Ocean (4.5° N to 5° S), a region previously not considered important for N_2 fixation[27], the depth-integrated N_2-fixation rates were on average 570% higher with the dissolution method (55 ± 7 μmol N m^{-2} d^{-1}) than with the bubble-addition method (8 ± 1 μmol N m^{-2} d^{-1}; means and propagated s.e.m., $n = 7$) (Fig. 3). Moreover, in the South Atlantic Gyre (38° S) and in the Falkland Current (44° S), where water temperatures were as low as 16° C and 10° C, respectively, considerable N_2 fixation was detected with the dissolution method (0.44 ± 0.1 and 0.54 ± 0.1 μmol N m^{-3} d^{-1}; means and s.e.m., $n = 3$). The rates detected with the bubble-addition method were substantially lower (0.10 ± 0.01 and 0.18 ± 0.03 μmol N m^{-3} d^{-1} for 38° S and 44° S, respectively; means and s.e.m., $n = 3$).

We used TaqMan assays based on the presence of the nitrogenase reductase (*nifH*) gene to quantify the relative abundance of the diazotrophic phylotypes found in the tropical Atlantic Ocean[20,21]. We identified two geographically separated areas, dominated by distinct diazotrophic phylotypes (Fig. 4 and Supplementary Fig. 2). The most abundant diazotroph was *Trichodesmium*, peaking at 9° N with 5 × 10^5 *Trichodesmium nifH* copies per litre. We measured the highest N_2-fixation rates inside a *Trichodesmium* bloom at 13.75° N with 360 ± 8.5 μmol N m^{-2} d^{-1} and 219 ± 8.2 μmol N m^{-2} d^{-1} measured with the dissolution and bubble-addition methods, respectively (means and s.e.m., $n = 3$). Although *Trichodesmium* was present throughout the whole area (Fig. 4), its abundance declined rapidly south of 5° N. Diatom-associated heterocystous diazotrophs, unicellular cyanobacteria and γ-proteobacteria showed a peak of abundance within the *Trichodesmium* bloom and a second peak in the equatorial Atlantic region, where they outnumbered *Trichodesmium*. We characterize the tropical North Atlantic (5° N to 15° N) as a region of *Trichodesmium* dominance, where underestimation of N_2-fixation rates by the bubble-addition method was less severe (but still significant). By contrast, we found the largest underestimation (570%) with the bubble-addition method in the equatorial Atlantic (4.5° N to 5° S), which was dominated by diazotrophs other than *Trichodesmium* (Supplementary Figs 4 and 5). The combined results indicate that the magnitude of underestimation in N_2-fixation rates measured with the bubble-addition method relative to the

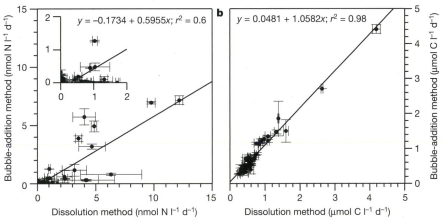

a
b
y = −0.1734 + 0.5955x; r² = 0.6
y = 0.0481 + 1.0582x; r² = 0.98
Bubble-addition method (nmol N l⁻¹ d⁻¹)
Dissolution method (nmol N l⁻¹ d⁻¹)
Bubble-addition method (μmol C l⁻¹ d⁻¹)
Dissolution method (μmol C l⁻¹ d⁻¹)

Figure 2 | Comparison between bubble-addition and dissolution methods. **a**, N_2-fixation rates for all stations and depths. Inset shows zoom on the 0–2-nmol l^{-1} d^{-1} range. **b**, Carbon-fixation rates for all stations and depths. Error bars indicate standard errors of triplicate incubations (s.e.m., $n = 3$).

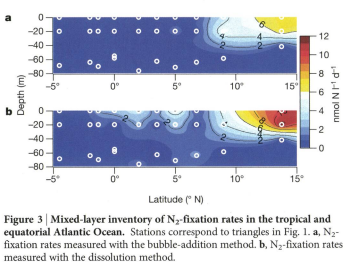


Figure 3 | Mixed-layer inventory of N_2-fixation rates in the tropical and equatorial Atlantic Ocean. Stations correspond to triangles in Fig. 1. **a**, N_2-fixation rates measured with the bubble-addition method. **b**, N_2-fixation rates measured with the dissolution method.

dissolution method was related to the composition of the diazotrophic community (Supplementary Fig. 2).

The species composition of the diazotrophic community can affect the level of underestimation for several reasons. In the bubble-addition method, the equilibration time for a $^{15}N_2$ gas bubble is longer than the typical incubation time of 24 h, despite manual shaking at the onset of the incubation[6] (Supplementary Fig. 6). This means that the bubble and its immediate surroundings will remain highly enriched in $^{15}N_2$, whereas water at the bottom of the incubation bottle will be less enriched (see Supplementary Information). As a result, buoyant diazotrophs (such as *Trichodesmium*) may be exposed to a higher fraction of the added $^{15}N_2$ label, whereas diazotrophs that sink (such as *Richelia*, when in symbiotic relationship with diatoms[28]) and accumulate at the bottom of the bottle may be exposed to a lower fraction of the $^{15}N_2$. Underestimation bias will also be greater when

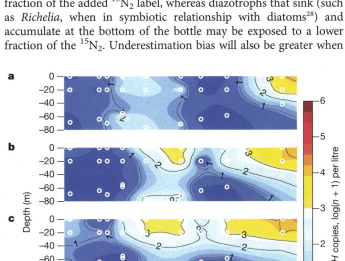


Figure 4 | Relative abundance of various phylotypes of diazotrophic bacteria from the same stations as the N_2-fixation rate measurements, estimated with TaqMan *nifH* gene assays. **a**, Unicellular cyanobacteria (Group A, Group B and Group C cyanobacteria). **b**, Diatom-associated heterocystous cyanobacteria. **c**, Diazotrophic γ-proteobacteria. **d**, *Trichodesmium*. There was no DNA sampling at the 3.5° N station. Note that the *nifH* gene copies (*n*) per litre of diatom-associated heterocystous cyanobacteria (such as *Richelia*) detects all cells of the symbionts, not just the heterocysts in which N_2 fixation is actively taking place.

the start of incubation coincides with the peak of N_2-fixation activity of the dominant diazotrophs (such as *Crocosphaera* for night-time incubations and *Trichodesmium* for daytime ones). These biases are avoided with the dissolution method because the $^{15}N_2$ label is added as a gas dissolved in sea water that is mixed uniformly with the sample at the start of the incubation.

Using published measurements of N_2-fixation rates obtained with the bubble-addition method, we calculated basin-wide and global budgets of marine N_2 fixation. The basin-wide averages are 14 Tg N yr^{-1} for the Atlantic, 63 Tg N yr^{-1} for the Pacific and 26 Tg N yr^{-1} for the Indian Ocean, giving a global total of 103 Tg N yr^{-1} (see Supplementary Information). Our own average N_2-fixation rates measured with the bubble-addition method and extrapolated over the Atlantic Ocean (25° N to 45° S), give an input of 14 Tg N yr^{-1}, which is identical to the value for the Atlantic Ocean calculated from the published data. By contrast, our average N_2-fixation rates measured with the dissolution method raise the Atlantic Ocean N_2-fixation rate to 24 Tg N yr^{-1}. If this relative difference applies equally to other ocean basins—where the diazotrophic communities may differ from those of the Atlantic Ocean—the global marine N_2-fixation rate based on direct measurements may have to be increased to 177 ± 8 Tg N yr^{-1}.

Our study demonstrates that field measurements of N_2 fixation made with the widely applied $^{15}N_2$ bubble-addition method have significantly and variably underestimated N_2-fixation rates in comparison with the dissolution method, with possible implications for the global marine nitrogen budget. Our data reveal regional variations in the magnitude of the underestimate, which may be related to the composition of the diazotrophic community. Specifically, the contribution of diazotrophs other than *Trichodesmium* may have been severely underestimated in previous field studies, leading to a biased view of the key players in this globally important process.

Received 3 November 2011; accepted 14 June 2012.
Published online 8 August 2012.

1. Karl, D. *et al.* Dinitrogen fixation in the world's oceans. *Biogeochemistry* **57,** 47–98 (2002).
2. Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281,** 200–206 (1998).
3. Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N. & Dunne, J. P. Spatial coupling of nitrogen inputs and losses in the ocean. *Nature* **445,** 163–167 (2007).
4. Altabet, M. A. Constraints on oceanic N balance/imbalance from sedimentary ^{15}N records. *Biogeosciences* **4,** 75–86 (2007).
5. Mahaffey, C., Michaels, A. F. & Capone, D. G. The conundrum of marine N_2 fixation. *Am. J. Sci.* **305,** 546–595 (2005).
6. Codispoti, L. A. An oceanic fixed nitrogen sink exceeding 400 Tg N a^{-1} vs the concept of homeostasis in the fixed-nitrogen inventory. *Biogeosciences* **4,** 233–253 (2007).
7. Montoya, J. P., Voss, M., Kahler, P. & Capone, D. G. A simple, high-precision, high-sensitivity tracer assay for N_2 fixation. *Appl. Environ. Microbiol.* **62,** 986–993 (1996).
8. Mohr, W., Grosskopf, T., Wallace, D. W. R. & LaRoche, J. Methodological underestimation of oceanic nitrogen fixation rates. *PLoS ONE* **5,** e12583 (2010).
9. Behrenfeld, M. J. & Falkowski, P. G. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol. Oceanogr.* **42,** 1–20 (1997).
10. Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. *Nature* **387,** 272–275 (1997).
11. Mills, M. M., Ridame, C., Davey, M., La Roche, J. & Geider, R. J. Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* **429,** 292–294 (2004).
12. Moore, J. K., Doney, S. C., Lindsay, K., Mahowald, N. & Michaels, A. F. Nitrogen fixation amplifies the ocean biogeochemical response to decadal timescale variations in mineral dust deposition. *Tellus B* **58,** 560–572 (2006).
13. Naqvi, S. W. A. *et al.* Increased marine production of N_2O due to intensifying anoxia on the Indian continental shelf. *Nature* **408,** 346–349 (2000).
14. Duce, R. A. *et al.* Impacts of atmospheric anthropogenic nitrogen on the open ocean. *Science* **320,** 893–897 (2008).
15. Gruber, N. & Galloway, J. N. An Earth-system perspective of the global nitrogen cycle. *Nature* **451,** 293–296 (2008).
16. Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11,** 235–266 (1997).
17. Zehr, J. P. *et al.* Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. *Nature* **412,** 635–638 (2001).
18. Zehr, J. P. *et al.* Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. *Science* **322,** 1110–1112 (2008).

19. Moisander, P. H. *et al.* Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. *Science* **327,** 1512–1514 (2010).
20. Langlois, R. J., LaRoche, J. & Raab, P. A. Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. *Appl. Environ. Microbiol.* **71,** 7910–7919 (2005).
21. Langlois, R. J., Hummer, D. & LaRoche, J. Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean. *Appl. Environ. Microbiol.* **74,** 1922–1931 (2008).
22. Hamersley, M. R. *et al.* Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight. *Aquat. Microb. Ecol.* **63,** 193–205 (2011).
23. Fernandez, C., Farias, L. & Ulloa, O. Nitrogen fixation in denitrified marine waters. *PLoS ONE* **6,** e20539 (2011).
24. Farnelid, H. *et al.* Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS ONE* **6,** e19223 (2011).
25. Halm, H. *et al.* Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *ISME J.* **6,** 1238–1249 (2012).
26. Luo, Y. W. *et al.* Database of diazotrophs in global ocean: abundances, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data Discuss.* **5,** 47–106 (2012).
27. Moore, C. M. *et al.* Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nature Geosci.* **2,** 867–871 (2009).
28. Subramaniam, A. *et al.* Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. *Proc. Natl Acad. Sci. USA* **105,** 10460–10465 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Klockgether and T. Max for mass-spectrometry measurements. We thank S. Fehsenfeld for helping with sampling and H. Nurlaeli for experimental work on *Nodularia*. We also thank the captain and crew of RV *Meteor* and RV *Polarstern*, as well as the chief scientists, P. Brandt and A. Macke. We thank D. Desai for statistical analyses. This work is a contribution of the Sonderforschungsbereich 754 'Climate — Biogeochemistry Interactions in the Tropical Ocean', which is supported by the Deutsche Forschungsgemeinschaft. We thank the Max Planck Gesellschaft for financial support. We thank the Bundesministerium für Bildung und Forschung (BMBF) for financial support through the SOPRAN II (Surface Ocean Processes in the Anthropocene) project, grant number 03F0611A.

Author Contributions W.M. designed the dissolution method. T.G., T.B., H.S. and D.G. collected samples and performed *nifH* gene quantification. M.M.M.K. and G.L. did the measurements on the mass spectrometer. T.G. wrote the manuscript with W.M. and J.L.R.. M.M.M.K., G.L., R.A.S., D.W.R.W., J.L.R., W.M. and T.G. designed the experiments and analysed the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.L.R. (julie.laroche@dal.ca) and T.G. (tgrosskopf@geomar.de).

LETTER

doi:10.1038/nature11358

More extreme swings of the South Pacific convergence zone due to greenhouse warming

Wenju Cai[1], Matthieu Lengaigne[2], Simon Borlace[1], Matthew Collins[3,4], Tim Cowan[1], Michael J. McPhaden[5], Axel Timmermann[6], Scott Power[7], Josephine Brown[7], Christophe Menkes[8], Arona Ngari[9], Emmanuel M. Vincent[2] & Matthew J. Widlansky[10]

The South Pacific convergence zone (SPCZ) is the Southern Hemisphere's most expansive and persistent rain band, extending from the equatorial western Pacific Ocean southeastward towards French Polynesia[1,2]. Owing to its strong rainfall gradient, a small displacement in the position of the SPCZ causes drastic changes to hydroclimatic conditions and the frequency of extreme weather events—such as droughts, floods and tropical cyclones—experienced by vulnerable island countries in the region[1–7]. The SPCZ position varies from its climatological mean location with the El Niño/Southern Oscillation (ENSO), moving a few degrees northward during moderate El Niño events and southward during La Niña events[2,5,6]. During strong El Niño events, however, the SPCZ undergoes an extreme swing—by up to ten degrees of latitude toward the Equator—and collapses to a more zonally oriented structure[5] with commensurately severe weather impacts[5,8–11]. Understanding changes in the characteristics of the SPCZ in a changing climate is therefore of broad scientific and socioeconomic interest. Here we present climate modelling evidence for a near doubling in the occurrences of zonal SPCZ events between the periods 1891–1990 and 1991–2090 in response to greenhouse warming, even in the absence of a consensus on how ENSO will change[12–14]. We estimate the increase in zonal SPCZ events from an aggregation of the climate models in the Coupled Model Intercomparison Project phases 3 and 5 (CMIP3[15] and CMIP5) multi-model database that are able to simulate such events. The change is caused by a projected enhanced equatorial warming in the Pacific[16] and may lead to more frequent occurrences of extreme events across the Pacific island nations most affected by zonal SPCZ events.

The SPCZ plays an important part in global circulation and is a major feature of the Southern Hemisphere's climate[1,2]. Its location largely controls rainfall, ocean circulation and tropical cyclogenesis patterns in the South Pacific[5–7]. The western, more equatorial portion of the SPCZ rainfall band is largely controlled by sea surface temperature (SST), whereas its eastern portion is also influenced by extratropical circulation and the subtropical dry zone of the southeastern Pacific[1,17]. As the SPCZ moves northward during El Niño events, countries located within the climatological SPCZ position experience forest fires and droughts[3,4] as well as an increased probability of tropical cyclone hits[5]. In addition, the associated environmental changes affect fisheries[18] and cause coral reef mortality through thermally induced coral bleaching[8–11] across the South Pacific. Observed zonal SPCZ events, characterized by a collapse of the meridional tilt of the rain band, often associated with an equatorward shift of the intertropical convergence zone, have occurred in conjunction with the strongest recorded El Niño events (for example, 1982/83, 1997/98), in which large SST anomalies develop in the central and eastern equatorial Pacific. The effects of these zonal SPCZ events are much more severe than those from weaker El Niño events, and include massive drought and food shortage[7], unprecedented coral bleaching-induced mortality[9,10] and cyclogenesis in the vicinity of French Polynesia[5], a region not accustomed to such occurrences.

These dramatic effects raise the question as to whether greenhouse warming will change the frequency of zonal SPCZ events. Although many studies have addressed the effects of a projected warming on the Pacific mean state[13,14,19,20], ENSO and its multiplicity[12–14,21,22] and the mean position of the SPCZ[23], the issue of how zonal SPCZ events will change in a warming climate has received little attention. Here we show that greenhouse warming leads to a significant increase in the frequency of such events.

We apply a multivariate signal processing method referred to as empirical orthogonal function (EOF) analysis to deconvolve the spatiotemporal rainfall variability into orthogonal modes, each described by a principal spatial pattern and an associated principal component time series[24]. For observations, we use the satellite-era rainfall data set[25] focusing on austral summer (December to February) when the SPCZ is best developed. The leading pattern (Fig. 1a) features opposite rainfall anomalies between the equatorial and southwestern Pacific around the climatological rain band position. The second pattern (Fig. 1b) is characterized by opposite rainfall anomalies in the equatorial western and central Pacific.

A nonlinear relationship exists between the two associated principal components (time series). In one cluster of events, the principal components are negatively correlated (blue, green and black dots in Fig. 1c) and rainfall anomalies embedded in the two patterns tend to offset each other east of the Date Line, resulting in rainfall variability located in the western Pacific. This cluster encompasses years characterized by the well-known northward and southward shift of the SPCZ during La Niña and moderate El Niño events, respectively (blue and green lines in Fig. 1a). In the second cluster, the principal components are both positive (red dots in Fig. 1c) and display a positive correlation, resulting in large precipitation anomalies east of the Date Line. This cluster consists of the three reported zonal SPCZ events[5] over the 1979–2011 period (1982/83, 1991/1992 and 1997/98, or one in 11 years) where the eastern portion moves equatorward by more than 10° in latitude (red line; Fig. 1b). A nonlinear relationship of the second principal component with a historical ENSO index (for example, Niño3.4) is apparent (Supplementary Fig. 1), although we do not distinguish the canonical from the Modoki[22] ENSO. We define a zonal SPCZ event as one for which the first principal component is greater than one standard deviation and the second principal component is greater than zero.

We use this nonlinear behaviour to assess the performance of 17 coupled general circulation models (CGCM) from the CMIP3[15] database forced with historical anthropogenic and natural forcings

[1]CSIRO Marine and Atmospheric Research, Aspendale, Victoria 3195, Australia. [2]Laboratoire d'Océanographie et du Climat: Expérimentation et Approches Numériques (LOCEAN), Tour 45-55 – 4ème étage – Pièce 408, Case 100 – UPMC, 4 Place Jussieu, 75252 Paris Cedex 05, France. [3]College of Engineering Mathematics and Physical Sciences, Harrison Building, Streatham Campus, University of Exeter, Exeter EX4 4QF, UK. [4]Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK. [5]NOAA/Pacific Marine Environmental Laboratory, Seattle, Washington 98115, USA. [6]IPRC, Department of Oceanography, SOEST, University of Hawaii, Honolulu, Hawaii 96822, USA. [7]Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Victoria 3001, Australia. [8]Institut de Recherche pour le Développement, 101 Promenade Roger Laroque, BP A5 – 98848, Noumea, New Caledonia. [9]Meteorological Service, PO Box 127, Avarua, Rarotonga, Cook Islands. [10]International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA.

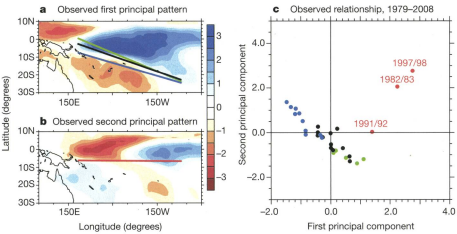


Figure 1 | Principal variability patterns of observed rainfall and their nonlinear relationship. **a**, **b**, Spatial patterns obtained by applying a statistical and signal processing method, EOF analysis[24], to a satellite-era rainfall anomaly data from the Global Precipitation Climatology Project version 2[25], focusing on the South Pacific domain (0°–30° S, 160° E–80° W) and the peak season for SPCZ, austral summer (December to February). The associated pattern beyond the domain is obtained by linearly regressing rainfall anomalies onto the time series. The first and second principal spatial patterns account for 47% and 16% of the total variance. The SPCZ position for El Niño (green line), La Niña (blue line) and neutral (black line) states is superimposed in **a**, and the position for zonal SPCZ events (red line) in **b**. The position of the SPCZ is defined as where the maximum rainfall is greater than 6 mm d^{-1} (ref. 5), and a linear fit is applied. Colour scale at right gives rainfall in mm d^{-1}; blue contours indicate increased rainfall, red contours indicate decreased rainfall per one standard deviation (s.d.) change. **c**, A nonlinear relationship between the associated principal component time series. La Niña, neutral and moderate El Niño years are indicated with blue, black and green dots, respectively. A zonal SPCZ event (red dots) is defined as when the first principal component is greater than one standard deviation, and when the second principal component is greater than zero. The different phases of ENSO events (El Niño and La Niña) are determined from a detrended Niño3.4 index when its amplitude is greater than 0.5 s.d.

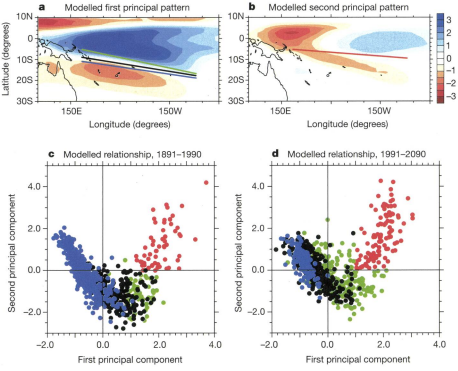


Figure 2 | Multi-model ensemble average of the principal variability patterns of rainfall and their nonlinear relationship from eight CMIP3 CGCMs that are able to produce the nonlinear relationship. **a**, **b**, The first (**a**) and second (**b**) principal variability patterns. The multi-model composite SPCZ position over the full 200 years of the control plus climate change periods (see below) for El Niño (green), La Niña (blue) and neutral (black) states is superimposed in **a**, and for zonal SPCZ events (red) in **b**. Colour scale at right gives rainfall in mm d^{-1}; blue contours indicate increased rainfall, and red contours indicate decreased rainfall per one s.d. change. **c**, **d**, The relationship between the associated two principal component time series for the control (1891–1990; **c**) and climate change (1991–2090; **d**) periods from the eight CMIP3 CGCMs. La Niña, neutral and moderate El Niño years are indicated with blue, black and green dots, respectively. Variance accounted for by the first and second principal pattern is model-dependent and ranges from 40–53% and 8–15% of the total variance, respectively. An increase in zonal SPCZ events (red dots) from the control to climate change period is evident from a comparison between **c** and **d**. An El Niño (or La Niña) is defined as when the detrended Niño3.4 index has an absolute value greater than 0.5 s.d.

before year 2000, and with a greenhouse gas emission scenario (SRESA2)[15] after year 2000 (see Methods). Nine models fail to generate the nonlinear behaviour (Supplementary Fig. 2). We focus on eight CGCMs that realistically simulate the two principal spatial patterns of rainfall anomalies (Fig. 2a and b), the nonlinear behaviour between the associated principal components (Fig. 2c and d), and the relationship with ENSO (Supplementary Figs 3 and 4). We compare the frequency of zonal SPCZ events in the first (1891–1990) and second (1991–2090) 100-year periods, which we refer to as the 'control' and the 'climate change' periods, respectively, to investigate the influence of greenhouse warming.

Aggregated over these eight CMIP3 CGCMs, the frequency of zonal SPCZ events increases by 81%, from about one event every 14 years in the control period to one every 7.5 years in the climate change period (Fig. 2c and d). This is statistically significant ($P < 0.001$) above the 99% confidence level based on a bootstrap test[26]. The statistical significance is underscored by a strong consensus among CGCMs, with six out of eight CGCMs simulating an increase in occurrences; a sensitivity test to varying definitions of zonal SPCZ events supports the robustness of our result (Supplementary Fig. 3 and Supplementary Table 1).

We conducted an identical analysis on rainfall outputs of 35 available experiments from 20 CMIP5 CGCMs under the historical emissions (before 2005) and a representative concentration pathway (RCP8.5, after 2005) scenario. Aggregated over 15 experiments from eight CMIP5 CGCMs that reproduce the nonlinear behaviour of the SPCZ, there is a similar increase in the occurrences of zonal SPCZ events, from one every 16 years in the control period to one every 7.5 years in the climate change period (Supplementary Table 2 and Supplementary Fig. 5).

We assess the potential impact of well-known biases in simulated climatological SSTs and SPCZ positions[23] using the HadCM3 CGCM (in which biases are corrected through a fixed flux adjustment[13]) in a set of 17 perturbed physics ensemble (PPE) climate change experiments forced with a 1% per year CO_2 increase. In these PPE experiments, perturbations are made to uncertain physical parameters within a single model structure. Aggregated over 12 experiments that produce the nonlinear behaviour of the SPCZ, there is a 214% increase in the frequency of zonal SPCZ events, from one event every 21 years in the control period to one event every 7 years in the climate change period (Supplementary Fig. 6 and Supplementary Table 3). Thus, the conclusions drawn from CMIP3 and CMIP5 CGCMs are not a consequence of the SST biases.

Understanding the relationship between the SPCZ and SST anomalies is crucial for unravelling the mechanism of the increase in zonal SPCZ events in a warming climate. To achieve this, modelled SST, wind and rainfall composite patterns are obtained for zonal SPCZ events (Fig. 3). In both periods, an eastward extension of the Pacific warm pool results in a substantial decrease of the meridional SST gradient south of the Equator, with a rainfall reduction over the western Pacific including northern Australia. In response, eastward-shifted westerly anomalies develop, with a maximum east of the Date Line and south of the Equator, accompanied by strong southeasterlies south of the warm SST anomalies (Fig. 3a and c), as in the observed SST and winds[5,27,28]. Such a wind response increases the moisture convergence east of the Date Line and south of the Equator, displacing the maximum moisture convergence and precipitation axis by up to 10° latitude to the north from its climatological position and contributing to the zonal orientation of the SPCZ (Fig. 3b and d). A zonal SPCZ event therefore occurs when the equatorial western Pacific moves from a state in which the off-equatorial region is warmer than the equatorial region to a condition in which the meridional SST gradient (the off-equatorial minus the equatorial) decreases considerably (Supplementary Figs 7 and 8). An increase in the frequency of zonal SPCZ events may hence arise from more frequent occurrences of a vanishing meridional SST gradient, which may be associated with a change in the occurrences of strong El Niño events. This mechanism also operates for the CMIP5 models for which data are available at the time of writing.

As shown in Supplementary Fig. 9, climate change (greenhouse warming) induces a larger warming rate along the Equator than in the off-equatorial central Pacific, a common feature of the response to global warming[16,29,30]. This pattern results in a slight decrease of the climatological meridional SST gradient in the central Pacific, which translates to an increase in the frequency of austral summers where the meridional SST gradient diminishes (for example, to a value <0.25 °C, Fig. 4a). Such a change occurs in all eight CMIP3 CGCMs (Fig. 4c), and in aggregation, the number of austral summers with a vanishing meridional SST gradient (that is, <0.25 °C) increases by 150%, from one every 15 summers during the control period to one every six summers during the climate change period. In association with this increase, for a given vanishing meridional SST gradient, the number of zonal SPCZ events increases (Fig. 4b).

In addition, there is a 32% increase in strong El Niño events (as measured by a Niño3.4 index greater than a 1.5-s.d. value; the Niño3.4

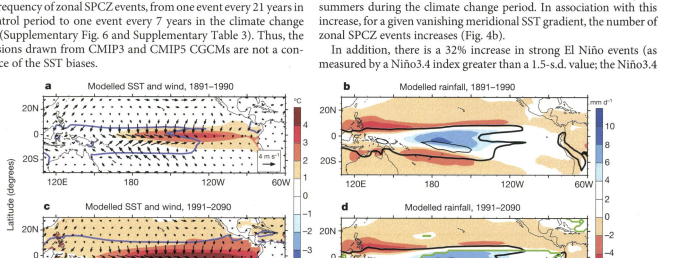


Figure 3 | Multi-model composites of the circulation fields associated with zonal SPCZ events. Shown are SST (°C), surface winds (m s^{-1}) and rainfall (mm d^{-1}) anomalies referenced to the mean climate of the control period. **a**, **b**, events over the control period; **c**, **d**, events over the climate change period. Contours of the composite 28 °C isotherm are superimposed on the SST anomalies (purple), and total rainfall of 6 mm d^{-1} (thick black contour) and 12 mm d^{-1} (thin black contour) are superimposed on the rainfall anomalies. The eight CMIP3 CGCM ensemble average shows an increase in rainfall intensity of ~10% averaged over the SPCZ area during such zonal SPCZ events, from the control period to the climate change period. In areas confined by the green curve, the increase is statistically significant ($P < 0.05$) above the 95% confidence level, based on a *t*-test comparing the two averages.

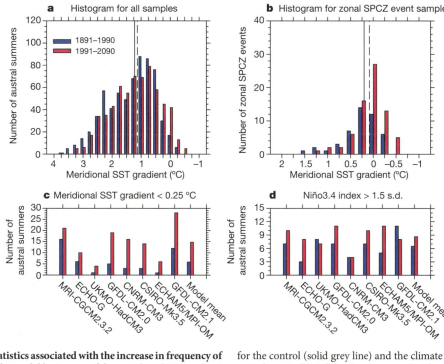


Figure 4 | Multi-model statistics associated with the increase in frequency of zonal SPCZ events. **a**, Multi-model histogram of the meridional SST gradient in the central equatorial Pacific for the eight CMIP3 CGCMs that are able to produce the nonlinear relationship for the control and climate change period. The meridional SST gradient is defined as the average SST over the off-equatorial region (10° S–5° S, 155° E–120° W) minus the average over the equatorial region (5° S-0°, 155° E–120° W). All 800 years in each period are separated in 0.25 °C bins centred at the tick point for the control (blue) and climate change (red) period. The multi-model median meridional SST gradient for the control (solid grey line) and the climate change (dashed grey line) periods are indicated. **b**, The same as **a** but for zonal SPCZ events only. Note that the horizontal axes in **a** and **b** are reversed; that is, decreasing go eastward. **c**, Changes in the number of austral summers in which the meridional SST difference virtually vanishes (that is, is <0.25 °C), with all eight CGCMs producing an increase. **d**, Changes in the number of strong El Niño events (the Niño3.4 index greater than a 1.5-s.d. value), with two CGCMs showing a decrease (UKMO-HadCM3 and GFDL-CM2.1).

index is defined as SST anomalies averaged over the region of 170° W–120° W, 5° S–5° N; Supplementary Table 4). This increase, seen in five out of eight CMIP3 CGCMs (Fig. 4d), occurs despite a tendency for a decrease in the total number of El Niño events (defined as when Niño3.4 is greater than a 0.5-s.d. or 0.75-s.d. value), and despite a general lack of consensus on how the El Niño amplitude may change[12,13]. However, only about 50% of the total increase in zonal SPCZ events coincides with a strong El Niño event, suggesting that such a change in El Niño is not necessary for an increase in zonal SPCZ events. Indeed, results from the 12 PPE experiments show a lack of consensus among the PPE experiments, despite a large increase in the frequency of zonal SPCZ events (Supplementary Fig. 10 and Supplementary Table 5), highlighting the fundamental importance of a decrease in the meridional SST gradient.

In summary, increased occurrences of zonal SPCZ events are consistent with the background state change in SST, and support our conclusion of an association with human-induced greenhouse warming, despite a lack of consensus on how ENSO will change. Although the simulated frequency in the control period is comparable to the observed frequency, the latter is based on observations over some three decades and may carry a large uncertainty. Thus, we cannot exclude the possibility that the observed frequency may be higher or lower if an extended observational record was available. With the projected large percentage increase, we expect more frequent occurrences of extreme events such as droughts, floods and tropical cyclones in the Pacific Island nations most affected by zonal SPCZ events.

METHODS SUMMARY

The zonal SPCZ events were diagnosed from observations and CGCMs. We propose an identification method based on the historical events, in which we apply EOF analysis[24] to the satellite-era austral summer rainfall anomalies for the 1979–2008 period over the South Pacific domain (160° E–80° W, 0°–30° S). The rainfall data[25], Global Precipitation Climatology Project monthly precipitation analysis, is obtained from http://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html. The rainfall anomalies are referenced to the climatological mean of the full period. This yields two leading patterns, one reflecting opposite rainfall anomalies between the equatorial and southwestern Pacific around the climatological rain band position, the other characterized by opposite rainfall anomalies in the equatorial western and central Pacific. The outputs of the analysis are arranged so that the principal component time series have a standard deviation of one, and the differences in variance are expressed in the pattern. Anomalies beyond the South Pacific domain are obtained by a linear regression. A zonal SPCZ event is defined as when the first leading time series is greater than one standard deviation, and when the second leading time series is greater than zero. This definition captures the three observed zonal SPCZ events in the second cluster. We also test the sensitivity of our results to varying definitions (Supplementary Table 1). The method is applied to rainfall anomalies of 17 CMIP3 CGCM simulations, each covering 110 years of a pre-twenty-first-century climate change simulation using historical anthropogenic and natural forcings (1891–2000) and another 90 years (2001–90; the longest common period for the CGCMs) from a future greenhouse warming simulation using the A2 greenhouse gas emission scenario (SRESA2)[15]. The same method is applied to 35 CMIP5 experiments under historical anthropogenic and natural forcings and the RCP8.5 forcing scenario, and 17 PPE experiments forced with historical and a 1% per year CO_2 increase, covering a 200-year period.

Full Methods and any associated references are available in the online version of the paper.

Received 11 April; accepted 27 June 2012.

1. Kiladis, G. N., Storch, H. V. & Loon, H. V. Origin of the South Pacific Convergence Zone. *J. Clim.* **2,** 1185–1195 (1989).
2. Vincent, D. G. The south Pacific convergence zone (SPCZ): a review. *Mon. Weath. Rev.* **122,** 1949–1970 (1994).
3. Salinger, M. J., Renwick, J. A. & Mullan, A. B. Interdecadal Pacific oscillation and South Pacific climate. *Int. J. Climatol.* **21,** 1705–1721 (2001).
4. Kumar, V. V., Deo, R. C. & Ramachandran, V. Total rain accumulation and rain-rate analysis for small tropical Pacific islands: a case study of Suva, Fiji. *Atmos. Sci. Lett.* **7,** 53–58 (2006).

5. Vincent, E. M. *et al.* Interannual variability of the South Pacific Convergence Zone and implications for tropical cyclone genesis. *Clim. Dyn.* **36,** 1881–1896 (2011).
6. Folland, C. K., Renwick, J. A., Salinger, M. J. & Mullan, A. B. Relative influence of the interdecadal Pacific oscillation and ENSO on the South Pacific Convergence Zone. *Geophys. Res. Lett.* **29,** 1643, http://dx.doi.org/10.1029/2001GL014201 (2002).
7. Hennessy, K. Power, S. & Cambers, G. (eds) *Climate Change in the Pacific: Scientific Assessment and New Research* Vol. 1, *Regional Overview*; Vol. 2, *Country Reports* (Australian Bureau of Meteorology and CSIRO, 2011).
8. Barnett, J. Dangerous climate change in Pacific Islands: food production and food security. *Reg. Environ. Change* **11,** 229–237 (2011).
9. Glynn, P. Widespread coral mortality and the 1982–83 El Niño warming event. *Environ. Conserv.* **11,** 133–146 (1984).
10. Hoegh-Guldberg, O. Climate change, coral beaching and the future of the world's coral reefs. *Mar. Freshwat. Res.* **50,** 839–866 (1999).
11. Mumby, P. J. *et al.* Unprecedented bleaching-induced mortality in *Porites* spp. at Rangiroa Atoll, French Polynesia. *Mar. Biol.* **139,** 183–189 (2001).
12. Guilyardi, E. *et al.* Understanding El Niño in ocean-atmosphere general circulation models: progress and challenges. *Bull. Am. Meteorol. Soc.* **90,** 325–340 (2009).
13. Collins, M. *et al.* The impact of global warming on the tropical Pacific Ocean and El Niño. *Nature Geosci.* **3,** 391–397 (2010).
14. Ashok, K., Sabin, T. P., Swapna, P. & Murtugudde, R. G. Is a global warming signature emerging in the tropical Pacific? *Geophys. Res. Lett.* **39,** L02701, http://dx.doi.org/10.1029/2011GL050232 (2012).
15. Meehl, G. *et al.* The WCRP CMIP3 multimodel dataset: a new era in climate change research. *Bull. Am. Meteorol. Soc.* **88,** 1383–1394 (2007).
16. Xie, S. P. *et al.* Global warming pattern formation: sea surface temperature and rainfall. *J. Clim.* **23,** 966–986 (2010).
17. Takahashi, K. & Battisti, D. S. Processes controlling the mean tropical Pacific precipitation pattern. Part II: the SPCZ and the southeast Pacific dry zone. *J. Clim.* **20,** 5696–5706 (2007).
18. Lehodey, P., Bertignac, M., Hampton, J., Lewis, A. & Picaut, J. El Niño Southern Oscillation and tuna in the western Pacific. *Nature* **389,** 715–718 (1997).
19. Cane, M. A. *et al.* Twentieth-century sea surface temperature trends. *Science* **275,** 957–960 (1997).
20. Vecchi, G. A. *et al.* Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature* **441,** 73–76 (2006).
21. Yeh, S.-W. *et al.* El Niño in a changing climate. *Nature* **461,** 511–514 (2009).
22. Ashok, K., Behera, S. K., Rao, S. A., Weng, H. & Yamagata, T. El Niño Modoki and its possible teleconnection. *J. Geophys. Res.* **112,** C11007, http://dx.doi.org/10.1029/2006JC003798 (2007).
23. Brown, J. R., Moise, A. F. & Delage, F. P. Changes in the South Pacific Convergence Zone in IPCC AR4 future climate projections. *Clim. Dyn.* **39,** 1–19 (2012).
24. Lorenz, E. N. *Empirical Orthogonal Functions and Statistical Weather Prediction* (Statistical Forecast Project Rep. 1, Department of Meteorology, MIT, 1956).
25. Adler, R. F. *et al.* The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeorol.* **4,** 1147–1167 (2003).
26. Austin, P. C. Bootstrap methods for developing predictive models. *Am. Stat.* **58,** 131–137 (2004).
27. Spencer, H. Role of the atmosphere in seasonal phase locking of El Niño. *Geophys. Res. Lett.* **31,** L24104, http://dx.doi.org/10.1029/2004GL021619 (2004).
28. Lengaigne, M., Boulanger, J. P., Menkes, C. & Spencer, H. Influence of the seasonal cycle on the termination of El Niño events in a coupled general circulation model. *J. Clim.* **19,** 1850–1868 (2006).
29. Liu, Z., Vavrus, S., He, F., Wen, N. & Zhong, Y. Rethinking tropical ocean response to global warming: the enhanced equatorial warming. *J. Clim.* **18,** 4684–4700 (2005).
30. Timmermann, A., McGregor, S. & Jin, F.-F. Wind effects on past and future regional sea level trends in the southern Indo-Pacific. *J. Clim.* **23,** 4429–4437 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Australian Climate Change Science Program, CSIRO Office of Chief Executive Science Leader programme, and the Pacific-Australia Climate Change Science and Adaptation Planning Program. A.T. and M.J.W. were supported by the Office of Science (BER) US Department of Energy, grant DE-FG02-07ER64469, the US National Science Foundation under grant 1049219 and by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). M.J.M. was supported by NOAA and by CSIRO as a visiting scholar. M.L., C.M. and E.M.V. were supported by the Institut de Recherche pour le Développement (IRD). This is PMEL contribution number 3830.

Author Contributions W.C. and M.L. conceived the study, directed the analysis and wrote the initial draft of the paper. S.B. performed the analysis. M.C. conducted the perturbed physics ensemble climate change experiments with the HadCM3 model. All authors contributed to interpreting results, discussion of the associated statistical significance, and improvement of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.C. (wenju.cai@csiro.au).

LETTER

doi:10.1038/nature11258

Reconstructing Native American population history

David Reich[1,2], Nick Patterson[2], Desmond Campbell[3,4], Arti Tandon[1,2], Stéphane Mazieres[3,5], Nicolas Ray[6], Maria V. Parra[3,7], Winston Rojas[3,7], Constanza Duque[3,7], Natalia Mesa[3,7], Luis F. García[7], Omar Triana[7], Silvia Blair[7], Amanda Maestre[7], Juan C. Dib[8], Claudio M. Bravi[3,9], Graciela Bailliet[9], Daniel Corach[10], Tábita Hünemeier[3,11], Maria Cátira Bortolini[11], Francisco M. Salzano[11], Maria Luiza Petzl-Erler[12], Víctor Acuña-Alonzo[13], Carlos Aguilar-Salinas[14], Samuel Canizales-Quinteros[15,16], Teresa Tusié-Luna[15], Laura Riba[15], Maricela Rodríguez-Cruz[17], Mardia Lopez-Alarcón[17], Ramón Coral-Vazquez[18], Thelma Canto-Cetina[19], Irma Silva-Zolezzi[20]†, Juan Carlos Fernandez-Lopez[20], Alejandra V. Contreras[20], Gerardo Jimenez-Sanchez[20]†, Maria José Gómez-Vázquez[21], Julio Molina[22], Ángel Carracedo[23], Antonio Salas[23], Carla Gallo[24], Giovanni Poletti[24], David B. Witonsky[25], Gorka Alkorta-Aranburu[25], Rem I. Sukernik[26], Ludmila Osipova[27], Sardana A. Fedorova[28], René Vasquez[29], Mercedes Villena[29], Claudia Moreau[30], Ramiro Barrantes[31], David Pauls[32], Laurent Excoffier[33,34], Gabriel Bedoya[7], Francisco Rothhammer[35], Jean-Michel Dugoujon[36], Georges Larrouy[36], William Klitz[37], Damian Labuda[30], Judith Kidd[38], Kenneth Kidd[38], Anna Di Rienzo[25], Nelson B. Freimer[39], Alkes L. Price[2,40] & Andrés Ruiz-Linares[3]

The peopling of the Americas has been the subject of extensive genetic, archaeological and linguistic research; however, central questions remain unresolved[1–5]. One contentious issue is whether the settlement occurred by means of a single[6–8] migration or multiple streams of migration from Siberia[2–15]. The pattern of dispersals within the Americas is also poorly understood. To address these questions at a higher resolution than was previously possible, we assembled data from 52 Native American and 17 Siberian groups genotyped at 364,470 single nucleotide polymorphisms. Here we show that Native Americans descend from at least three streams of Asian gene flow. Most descend entirely from a single ancestral population that we call 'First American'. However, speakers of Eskimo–Aleut languages from the Arctic inherit almost half their ancestry from a second stream of Asian gene flow, and the Na-Dene-speaking Chipewyan from Canada inherit roughly one-tenth of their ancestry from a third stream. We show that the initial peopling followed a southward expansion facilitated by the coast, with sequential population splits and little gene flow after divergence, especially in South America. A major exception is in Chibchan speakers on both sides of the Panama isthmus, who have ancestry from both North and South America.

The settlement of the Americas occurred at least 15,000 years ago through Beringia, a land bridge between Asia and America that existed during the ice ages[1–5]. Most analyses of Native American genetic diversity have examined single loci, particularly mitochondrial DNA or the Y chromosome, and some interpretations of these data model the settlement of America as a single migratory wave from Asia[6–8]. We assembled native population samples from Canada to the southern tip of South America, genotyped them on single nucleotide polymorphism (SNP) microarrays, and merged our data with six other data sets. The combined data set consists of 364,470 SNPs genotyped in 52 Native American populations (493 samples; Fig. 1a and Supplementary Table 1), 17 Siberian populations (245 samples; Supplementary Fig. 1 and Supplementary Table 2) and 57 other populations (1,613 samples) (Supplementary Notes).

A complication in studying Native American genetic history is admixture with European and African immigrants since 1492. Cluster analysis[16] shows that many of the samples we examined have some non-native admixture (an average of 8.5%; Fig. 1b and Supplementary Tables 1 and 3). This admixture is a challenge for learning about the historical relationships among the populations, and to address this complication we used three independent approaches. First, we restricted analyses to 163 Native Americans from 34 populations without evidence of admixture (Supplementary Notes). Second, we subtracted the expected contribution of European and African ancestry to the statistics we used to learn about population relationships (Supplementary Notes). Third, we inferred the probability of non-native ancestry at each genomic segment and 'masked' segments with more than a negligible probability of this ancestry (Fig. 1b,

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [3]Department of Genetics, Evolution and Environment, University College London WC1E 6BT, UK. [4]Department of Psychiatry and Centre for Genomic Sciences, The University of Hong Kong, Pokfulam, Hong Kong SAR. [5]Anthropologie Bio-culturelle, Droit, Ethique et Santé (ADES), UMR 7268, Aix-Marseille Université/CNRS/EFS, Marseille 13344, France. [6]Institute for Environmental Sciences, and Forel Institute, University of Geneva, Geneva 1227, Switzerland. [7]Universidad de Antioquia, Medellín, Colombia. [8]Fundación Salud para el Trópico, Santa Marta, Colombia. [9]Instituto Multidisciplinario de Biología Celular (CCT La Plata-CONICET, CICPCA), 1900 La Plata, Argentina. [10]Servicio de Huellas Digitales Genéticas and CONICET, Universidad de Buenos Aires, Argentina. [11]Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre 91501-970, Brazil. [12]Departamento de Genética, Universidade Federal do Paraná, Curitiba 81531-980, Brazil. [13]National Institute of Anthropology and History, México City 06100, México. [14]Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, México City 14100, México. [15]Unidad de Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán/Universidad Nacional Autónoma de México, México City 14000, México. [16]Departamento de Biología, Facultad de Química, Universidad Nacional Autónoma de México, México City 04510, México. [17]Unidad de Investigación Médica en Nutrición, Hospital de Pediatría, CMNSXXI, Instituto Mexicano del Seguro Social, México City 06720, México. [18]Sección de Posgrado, Escuela Superior de Medicina del Instituto Politécnico Nacional, México City 11340, México. [19]Laboratorio de Biología de la Reproducción, Departamento de Salud Reproductiva y Genética, Centro de Investigaciones Regionales, Mérida Yucatán 97000, México. [20]Instituto Nacional de Medicina Genómica, México City 14610, México. [21]Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Nuevo León 66451, México. [22]Centro de Investigaciones Biomédicas de Guatemala, Ciudad de Guatemala, Guatemala. [23]Instituto de Ciencias Forenses, Universidade de Santiago de Compostela, Fundación de Medicina Xenómica (SERGAS), CIBERER, Santiago de Compostela, Galicia 15782, Spain. [24]Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima 15102, Peru. [25]Department of Human Genetics, University of Chicago, Chicago 60637, USA. [26]Laboratory of Human Molecular Genetics, Institute of Molecular and Cellular Biology, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. [27]Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. [28]Department of Molecular Genetics, Yakut Research Center of Complex Medical Problems and North-East Federal University, Yakutsk, Sakha (Yakutia) 677010, Russia. [29]Instituto Boliviano de Biología de la Altura, Universidad Autonoma Tomás Frías, Potosí, Bolivia. [30]Département de Pédiatrie, Centre de Recherche du CHU Sainte-Justine, Université de Montréal, Montréal, Quebec H3T 1C5, Canada. [31]Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica. [32]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. [33]Computational and Molecular Population Genetics Laboratory, Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland. [34]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [35]Instituto de Alta Investigación, Universidad de Tarapacá, Programa de Genética Humana ICBM Facultad de Medicina Universidad de Chile and Centro de Investigaciones del Hombre en el Desierto, Arica 1001236, Chile. [36]Anthropologie Moléculaire et Imagerie de Synthèse, CNRS UMR 5288, Université Paul Sabatier Toulouse III, Toulouse 31000, France. [37]School of Public Health, University of California, Berkeley, California 94720, USA. [38]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [39]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, California 90095, USA. [40]Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. †Present addresses: BioAnalytical Science Department Nestec Ltd, Nestlé Research Center, 1000 Lausanne, Switzerland (I.S.-Z.); Global Biotech Consulting Group, México City 09010, México (G.J.-S.).

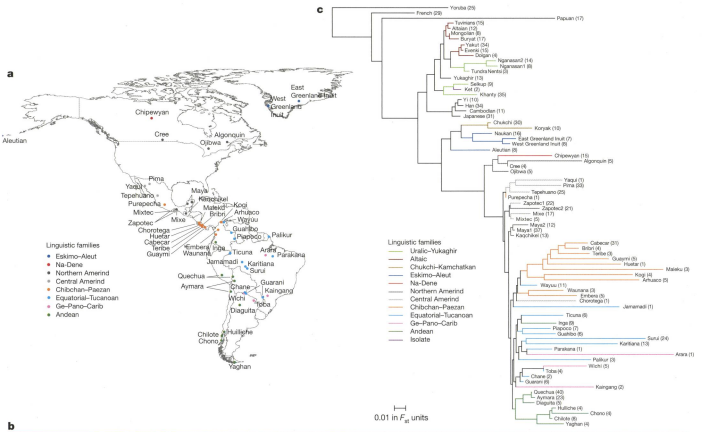


Figure 1 | Geographic, linguistic and genetic overview of 52 Native American populations. a, Sampling locations of the populations, with colours corresponding to linguistic groups. **b**, Cluster-based analysis ($k = 4$) using ADMIXTURE shows evidence of some West-Eurasian-related and sub-Saharan-African-related ancestry in many Native Americans before masking (top), but little afterwards (bottom). Thick vertical lines denote major linguistic groupings, and thin vertical lines separate individual populations. **c**, Neighbour-joining tree based on F_{st} distances relating Native American to selected non-American populations (sample sizes in parentheses). Native American and Siberian data were analysed after masking, but consistent trees were obtained on a subset of completely unadmixed samples (Supplementary Fig. 3). Some populations have evidence for substructure, and we represent these as two different groups (for example Maya1 and Maya2).

Supplementary Notes and Supplementary Fig. 2). Our inferences from these three approaches are concordant (Supplementary Figs 3 and 4).

We built a tree (Fig. 1c) using F_{st} distances between pairs of populations, which broadly agrees with geography and linguistic categories[17] (trees based on masked and unmasked data were similar; Supplementary Fig. 3). An early split separates Asians from Native Americans and extreme northeastern Siberians (Chukchi, Naukan, Koryak), which is consistent with studies that have identified pan-American variants shared with some northeastern Siberians[6,7,10,18]. Eskimo–Aleut speakers and far-northeastern Siberians form a cluster that is separated from other Native American populations by a long internal branch. Within America the tree shows a series of splits in an approximate north–south sequence beginning with the Arctic, followed by northern North America, northern/central and southern Mexico and lower Central America/Colombia, and ending in three South American clusters (the Andes, the Chaco region and eastern South America). This pattern of splits is consistent with a north–south population expansion, an inference that is also supported by the negative correlation between heterozygosity and distance from the Bering Strait ($r = -0.48$, $P = 0.007$). This correlation increases if we use 'least cost distances' that consider the coasts as facilitators of migration[19–21], and persists if we exclude four Native North American populations with ancestry from later streams of Asian gene flow (Supplementary Notes and Supplementary Fig. 5).

Trees provide a simplified model of history that does not accommodate the possibility of gene flow after population separation. Circumstantial evidence that some Native American populations may not fit a simple tree comes from cluster analysis, which infers Siberian-related ancestry in some northern North Americans (Fig. 1b), and from single-locus studies that have identified genetic variants shared between Eurasia and North America that are absent from South America[11,22,23]. The advent of genome-wide data sets has allowed the development of a formal four-population test for whether sets of four populations are consistent with a tree. This test is robust to the

Table 1 | Native Americans descend from at least three streams of Asian gene flow

Population groupings tested	*P* value for this many Asian streams being enough to explain the data			Minimum number of streams of Asian gene flow needed to explain the data
	1	2	3	
East Greenland Inuit/West Greenland Inuit/First American	$<10^{-9}$	0.64	1	2
East Greenland Inuit/Aleutian/First American	$<10^{-9}$	0.57	1	2
West Greenland Inuit/Aleutian/First American	$<10^{-9}$	0.41	1	2
Chipewyan/East Greenland Inuit/First American	$<10^{-9}$	0.02	1	3
Chipewyan/West Greenland Inuit/First American	$<10^{-9}$	0.006	1	3
Chipewyan/Aleutian/First American	$<10^{-9}$	0.03	1	3
Saqqaq/East Greenland Inuit/First American	$<10^{-9}$	6×10^{-6}	1	3
Saqqaq/West Greenland Inuit/First American	$<10^{-9}$	2×10^{-6}	1	3
Saqqaq/Aleutian/First American	$<10^{-9}$	0.17	1	2
Saqqaq/Chipewyan/First American	$<10^{-9}$	0.29	1	2
Saqqaq/Eskimo–Aleut/Chipewyan/First American	$<10^{-9}$	8×10^{-6}	0.27	3

We use the method described in Supplementary Notes to test formally whether specified groupings of Native American populations are consistent with descending from one, two or three streams of gene flow from Asia. We use 'First American' to refer to a pool of 43 populations from Meso-America southward, and 'Eskimo–Aleut' to refer to a pool of East and West Greenland Inuit and Aleuts. We test either three or four population groupings (when there are three groupings, the maximum number of streams we can reject is two, and so the *P* value for three streams is always 1). At least two streams of Asian gene flow are required to explain all rows ($P<10^{-5}$). The Chipewyan, Eskimo–Aleut and First Americans can only be jointly explained by at least three streams. Analysis of the Saqqaq Palaeo-Eskimo (using about sixfold fewer SNPs than for the other analyses) show that the Asian ancestry in this individual has a component that is different from that in First Americans and Greenland Inuit, but indistinguishable from the Chipewyan.

ascertainment bias affecting SNP arrays[24]. For each of the 52 Native American populations in turn, we tested the hypothesis that they conform to the tree: ((test population, southern Native American), (outgroup1, outgroup2)) for 45 pairs of ten Asian outgroups. We used a Hotelling *T*-test to evaluate whether all four-population test f_4 statistics of this form are consistent with the expectation of zero (Supplementary Notes). The test is not significant for 47 populations, which is consistent with their stemming from the same, presumably first, wave of American settlement; we call this ancestry 'First American' (Table 1). In contrast, four populations from northern North America show highly significant evidence of ancestry from additional streams of gene flow from Asia, subsequent to the initial peopling of America, which we confirm through the Hotelling *T*-test and a complementary test (Supplementary Notes): East Greenland Inuit ($P<10^{-9}$), West Greenland Inuit ($P<10^{-9}$), Aleutian Islanders ($P=9\times10^{-5}$) and Chipewyan ($P<10^{-9}$). The recently sequenced genome of a 4,000-year-old Saqqaq Palaeo-Eskimo from Greenland[25] also has evidence of ancestry that is distinct from more southern Native Americans ($P=2\times10^{-9}$) (Supplementary Notes).

Examination of the values of the f_4 statistics allows us to infer the minimum number of gene flow events from Asia into America consistent with the data. Each stream of gene flow is expected to produce a distinct vector of f_4 statistics, constituting a 'signature' of how the ancestral migrating population relates to present-day Asian populations. By finding the minimum number of vectors whose linear combinations are necessary to produce the vector observed in each population, we infer that a minimum of three gene flow events from Asia are necessary to explain the data from all Native American populations jointly, including the Saqqaq Palaeo-Eskimo (Supplementary Notes). These three episodes correspond to First American ancestry (distributed throughout the Americas) and to two additional streams of gene flow detected in a subset of northern North Americans (East Greenland Inuit, West Greenland Inuit, Aleutian Islanders, Chipewyan and Saqqaq). Table 1 shows that f_4 statistics in the Inuit and Aleutian islanders are consistent with deriving the non-First-American portions of their ancestry from the same later stream of Asian gene flow, providing support for deep shared ancestry between these linguistically linked groups[12,26]. The Na-Dene-speaking Chipewyan have a different pattern of f_4 statistics from Eskimo–Aleut speakers, implying that they descend at least in part from a separate stream of Asian gene flow ($P<10^{-9}$ for comparisons with the Greenland Inuit; Table 1). This is consistent with the hypothesis that Na-Dene languages mark a distinct migration from Asia[9,17]. Because we only have data from one Na-Dene-speaking group, an important direction for future work will be to test whether the distinct Asian ancestry that we detect in the Chipewyan is a shared signature throughout Na-Dene speakers. Finally, the Saqqaq[25] have a vector of f_4 statistics consistent with that in the Chipewyan, raising the possibility that the Saqqaq and Chipewyan both carry genetic material from the same later stream of Asian gene flow into the Americas, postdating the First American migration (Supplementary Notes).

To develop an explicit model for the settlement of the Americas, we used the admixture graph (AG) framework[24]. AGs are generalizations of trees that accommodate the possibility of a limited number of unidirectional gene flow events. They are powerful tools for learning about history because they make predictions about the values of *f*-statistics (such as f_4) that can be used to test the fit of a proposed model[24] (Supplementary Notes). Figure 2 presents an AG relating selected Native American and Old World populations that is a good fit to the data in the sense that none of the *f*-statistics predicted by the

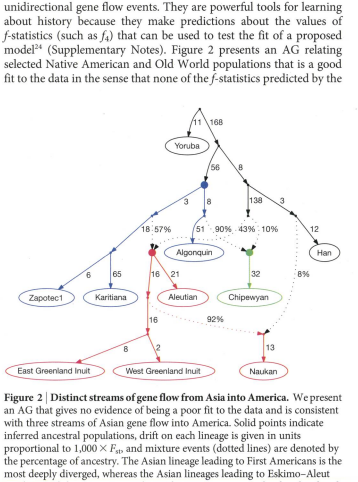


Figure 2 | Distinct streams of gene flow from Asia into America. We present an AG that gives no evidence of being a poor fit to the data and is consistent with three streams of Asian gene flow into America. Solid points indicate inferred ancestral populations, drift on each lineage is given in units proportional to $1{,}000\times F_2$, and mixture events (dotted lines) are denoted by the percentage of ancestry. The Asian lineage leading to First Americans is the most deeply diverged, whereas the Asian lineages leading to Eskimo–Aleut speakers and the Na-Dene-speaking Chipewyan are more closely related and descend from a common Siberian ancestral population that is a sister group to the Han. The inferred ancestral populations are indicated by filled circles, and the lineages descending from them are coloured: First American (blue), ancestors of the Na-Dene-speaking Chipewyan (green), and Eskimo–Aleut (red). The model also infers a migration of people related to Eskimo–Aleut speakers across the Bering Strait, thus bringing First American genes to Asia (the Naukan are shown, but the Chukchi show a similar pattern; Supplementary Notes).

model are more than three standard errors from what is observed. This supports the hypothesis of three deep lineages in Native Americans: the Asian lineage leading to First Americans is the most deeply diverged, whereas the Asian lineages leading to Eskimo–Aleut speakers and the Na-Dene-speaking Chipewyan are more closely related and descend from a putative Siberian ancestral population more closely related to Han (Fig. 2). We also arrive at the finding that Eskimo–Aleut populations and the Chipewyan derive large proportions of their genomes from First American ancestors: an estimated 57% for Eskimo–Aleut speakers, and 90% in the Chipewyan, probably reflecting major admixture events of the two later streams of Asian migration with the First Americans that they encountered after they arrived (Supplementary Notes). The high proportion of First American ancestry explains why Eskimo–Aleut and Chipewyan populations cluster with First Americans in trees like that in Fig. 1c despite having some of their ancestry from later streams of Asian migration, and explains the observation of some genetic variants that are shared by all Native Americans but are absent elsewhere[6,7,10,18]. We also infer back-migration of populations related to the Eskimo–Aleut from America into far-northeastern Siberia (we obtain an excellent fit to the data when we model the Naukan and coastal Chukchi as mixtures of groups related to the Greenland Inuit and Asians (Fig. 2 and Supplementary Notes)). This explains previous findings of pan-American alleles also in far-northeastern Siberia[6,7,10,18].

We next used AGs to develop a model for the history of populations who derive all their ancestry from the First American migration, with no ancestry from subsequent streams of Asian gene flow. Figure 3 presents an AG we built for 16 selected Native American populations and two outgroups, which is a good fit to the data in that the largest $|Z|$-score for a difference between the observed and predicted f-statistics is 3.2 from among the 11,781 statistics we tested (Supplementary Notes) (The AG of Fig. 3 used masked data; however, a consistent set of relationships is inferred for unadmixed samples (Supplementary Fig. 4).) This model provides a greatly improved statistical fit to the data compared with the tree of Fig. 1c and leads to several novel inferences. First, a relatively large fraction of South American populations fit the AG without a need for admixture events, which we speculate reflects a history of limited gene flow among these populations since their initial divergence. In contrast, only a small fraction of Meso-American populations fit into the AG, which could reflect either a higher rate of migration among neighbouring groups or our denser sampling in Meso-America allowing us to detect more subtle gene flow events. Second, some Meso-American populations have experienced very little genetic drift since divergence from the common ancestral population with South Americans (adding up the genetic drifts along the relevant edges of Fig. 3, we infer $F_{st} = 0.014$ between the Zapotec and a hypothetical population ancestral to all of Central and South America), suggesting that effective population sizes in Meso-America have been relatively large since settlement of the region. Third, the model infers three admixture events consistent with geographic locations and linguistic affiliations (Supplementary Notes). The Inga have both Amazonian and Andean ancestry, which is consistent with their speaking a Quechuan language but living in the eastern Andean slopes of Colombia and thus interacting with groups in the neighbouring Amazonian lowlands. The Guarani stem from two distinct strands of ancestry within eastern South America. The most striking admixture event is in the Costa Rican Cabecar (Fig. 3) and other Chibchan-speaking populations (Supplementary Notes) from the Isthmo-Colombian area. One of the lineages that we detect in these populations occurs definitively within the radiation of South American populations, and so the presence of these populations in lower Central America suggests that there was reverse gene flow across the Panama isthmus after the initial settlement of South America. There has been controversy about whether Chibchan speakers of lower Central America represent direct descendants of the first settlers in the region or more recent migration across the isthmus, and our results support the view that more recent migration has contributed most of these populations' ancestry[27].

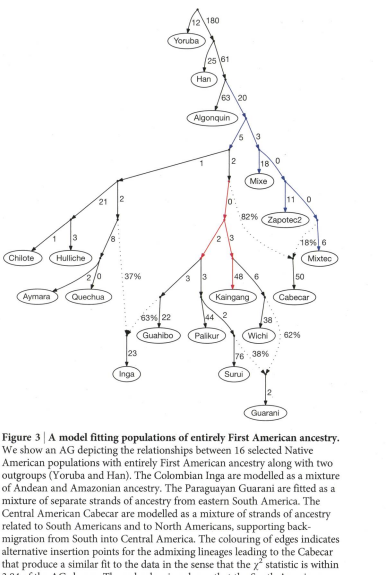


Figure 3 | A model fitting populations of entirely First American ancestry. We show an AG depicting the relationships between 16 selected Native American populations with entirely First American ancestry along with two outgroups (Yoruba and Han). The Colombian Inga are modelled as a mixture of Andean and Amazonian ancestry. The Paraguayan Guarani are fitted as a mixture of separate strands of ancestry from eastern South America. The Central American Cabecar are modelled as a mixture of strands of ancestry related to South Americans and to North Americans, supporting back-migration from South into Central America. The colouring of edges indicates alternative insertion points for the admixing lineages leading to the Cabecar that produce a similar fit to the data in the sense that the f^2 statistic is within 3.84 of the AG shown. The red colouring shows that the South American lineage contributing to the Cabecar split off after the divergence of the Andean populations, and the blue colouring shows that the other lineage present in the Cabecar diverged before the separation of Andeans. Estimated admixture proportions are shown along the dotted lines, and lineage-specific drift estimates are in units proportional to $1{,}000 \times F_{st}$.

This is the most comprehensive survey of genetic diversity in Native Americans so far. Our analyses show that the great majority of Native American populations—from Canada to the southern tip of Chile—derive their ancestry from a homogeneous 'First American' ancestral population, presumably the one that crossed the Bering Strait more than 15,000 years ago[6–8]. We also document at least two additional streams of Asian gene flow into America, allowing us to reject the view that all present-day Native Americans stem from a single migration wave[6–8], and supporting the more complex scenarios proposed by some other studies[9–15]. In particular, the three distinct Asian lineages we detect—'First American', 'Eskimo–Aleut' and a separate one in the Na-Dene-speaking Chipewyan—are consistent with a three-wave model proposed[9] mostly on the basis of dental morphology and a controversial interpretation of the linguistic data. However, our analyses also document extensive admixture between First Americans and the subsequent streams of Asian migrants, which was not predicted by that model, such that Eskimo–Aleut speakers and the Chipewyan

derive more than half their ancestry from First Americans. Further insights into Native American history will benefit from the application of analyses similar to those performed here to whole-genome sequences and to data from the many admixed populations in the Americas that do not self-identify as native[28–30].

METHODS SUMMARY

The DNA samples we analysed were collected over several decades. For each sample we verified that informed consent was obtained consistent with studies of population history and that institutional approval had been obtained in the country of collection. Ethical oversight and approval for this project was provided by the National Health Service National Research Ethics Service, Central London committee (reference no. 05/Q0505/31). The data set is based on merging Illumina SNP array data newly generated for this study (including 273 Native American samples) with data from six other studies. We applied stringent data curation and validation procedures to the merged data set. We used local ancestry inference software to identify genome segments in each Native American and Siberian sample without evidence of recent European or African admixture, and created a data set that masked segments of potentially non-native origin. Most analyses are performed on the masked data set; however, we confirmed major inferences on a subset of 163 Native American samples that had no evidence of European or African admixture. We used model-based clustering and neighbour-joining trees to obtain an overview of population relationships, and then tested whether proposed sets of four populations were consistent with having a simple tree relationship using the four-population test, which we generalized by means of a Hotelling T-test. We analysed the correlation in allele frequency differences across populations to infer the minimum number of gene flow events that occurred between Asia and America. We fitted the patterns of correlation in allele frequency differences to proposed models of history—AGs—that can incorporate population splits and mixtures.

Full Methods and any associated references are available in the online version of the paper.

Received 1 September 2011; accepted 25 May 2012.
Published online 11 July 2012.

1. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, 1994).
2. Meltzer, D. J. *First Peoples in a New World: Colonizing Ice Age America* (Univ. of California Press, 2009).
3. Goebel, T., Waters, M. R. & O'Rourke, D. H. The late Pleistocene dispersal of modern humans in the Americas. *Science* **319,** 1497–1502 (2008).
4. Dillehay, T. D. Probing deeper into first American studies. *Proc. Natl Acad. Sci. USA* **106,** 971–978 (2009).
5. O'Rourke, D. H. & Raff, J. A. The human genetic history of the Americas: the final frontier. *Curr. Biol.* **20,** R202–R207 (2010).
6. Tamm, E. *et al.* Beringian standstill and spread of Native American founders. *PLoS ONE* **2,** e829 (2007).
7. Kitchen, A., Miyamoto, M. M. & Mulligan, C. J. A three-stage colonization model for the peopling of the Americas. *PLoS ONE* **3,** e1596 (2008).
8. Fagundes, N. J. *et al.* Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* **82,** 583–592 (2008).
9. Greenberg, J. H., Turner, C. G. & Zegura, S. L. The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence. *Curr. Anthropol.* **27,** 477–497 (1986).
10. Leli, J. T. *et al.* The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* **70,** 192–206 (2002).
11. Bortolini, M. C. *et al.* Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet.* **73,** 524–539 (2003).
12. Volodko, N. V. *et al.* Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *Am. J. Hum. Genet.* **82,** 1084–1100 (2008).
13. Ray, N. *et al.* A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol. Biol. Evol.* **27,** 337–345 (2010).
14. de Azevedo, S. *et al.* Evaluating microevolutionary models for the early settlement of the New World: the importance of recurrent gene flow with Asia. *Am. J. Phys. Anthropol.* **146,** 539–552 (2011).
15. Perego, U. A. *et al.* Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* **19,** 1–8 (2009).
16. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
17. Ruhlen, M. *A Guide to the World's Languages* (Stanford Univ. Press, 1991).
18. Schroeder, K. B. *et al.* A private allele ubiquitous in the Americas. *Biol. Lett.* **3,** 218–223 (2007).
19. Ray, N. PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Mol. Ecol. Notes* **5,** 177–180 (2005).
20. Wang, S. *et al.* Genetic variation and population structure in native Americans. *PLoS Genet.* **3,** e185 (2007).
21. Yang, N. N. *et al.* Contrasting patterns of nuclear and mtDNA diversity in Native American populations. *Ann. Hum. Genet.* **74,** 525–538 (2010).
22. Brown, M. D. *et al.* mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am. J. Hum. Genet.* **63,** 1852–1861 (1998).
23. Karafet, T. M. *et al.* Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64,** 817–831 (1999).
24. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489–494 (2009).
25. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757–762 (2010).
26. Balter, M. Archaeology. The peopling of the Aleutians. *Science* **335,** 158–161 (2012).
27. Cooke, R. Prehistory of native Americans on the Central American land bridge: Colonization, dispersal, and divergence. *J. Archaeol. Res.* **13,** 129–187 (2005).
28. Wang, S. *et al.* Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet.* **4,** e1000037 (2008).
29. Bryc, K. *et al.* Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl Acad. Sci. USA* **107** (Suppl 2), 8954–8961 (2010).
30. Wall, J. D. *et al.* Genetic variation in Native Americans, inferred from Latino SNP and resequencing data. *Mol. Biol. Evol.* **28,** 2231–2237 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the volunteers who provided the samples that made this study possible. We thank E. D. Ruiz for assistance in the collection involving the Mixtec, Zapotec and Mixe; and P. Herrera for assistance in the collection involving the Quechua; A. Carnevale, M. Crawford, M. Metspalu, F. C. Nielsen, X. Soberon, R. Villems and E. Willerslev for facilitating sharing of data from Mexican, Siberian and Arctic populations; C. Stevens and A. Crenshaw for assistance with genotyping; and P. Bellwood, D. Bolnick, K. Bryc, J. Diamond, T. Dillehay, R. Gonzalez-José, M. Hammer, J. Hill, B. Kemp, S. LeBlanc, D. Meltzer, P. Moorjani, A. Moreno-Estrada, B. Pakendorf, J. Pickrell, M. Ruhlen, D. G. Smith, M. Stoneking, N. Tuross and A. Williams for critiques and discussions. Support was provided by National Institutes of Health grants NS043538 (A.R.-L.), NS037484 and MH075007 (N.B.F.), GM079558 (A.D.), GM079558-S1 (A.D.), GM057672 (K.K.K. and J.R.K.), and HG006399 (D.R., N.P. & A.L.P.); by a Biotechnology and Biological Sciences Research Council grant BB/I021213/1; by a National Science Foundation HOMINID grant BCS-1032255 (D.R. and N.P.); by a Canadian Institutes of Health Research grant (D.L.); by a Universidad de Antioquia CODI grant (G.B.); by a Fondo de Investigación Sanitaria grant PS 09/2368 (A.C.); by a Ministerio de Ciencia e Innovacion grant SAF2011-26983 (A.S.); by a Wenner-Gren Foundation grant ICRG-65 (A.D. and R.S.); by Russian Foundation for Basic Research grants 06-04-048182 (R.S.) and 02-06-80524a (L.O.); by a Siberian Branch Russian Academy of Sciences field grant (L.O.); by a Centre National de la Recherche Scientifique Programme Interdisciplinaire de Recherche Amazonie grant (J.-M.D.); and by startup funds from Harvard Medical School (D.R.) and the Harvard School of Public Health (A.L.P.).

Author Contributions D.R., N.B.F., A.L.P. and A.R.-L. conceived the project. D.R., N.P., D.C., A.T., S.M., N.R. and A.R.-L. performed analyses. D.R. and A. R.-L. wrote the paper with input from all the co-authors. A.R.-L. assembled the sample collection, directed experimental work and coordinated the study. All other authors contributed to the collection of samples and data.

Author Information The data analysed here are available for non-profit research on population history under an inter-institutional data access agreement with the Universidad de Antioquia, Colombia; queries regarding data access should be sent to A.R.-L. (a.ruizlin@ucl.ac.uk). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu) or A.R.-L. (a.ruizlin@ucl.ac.uk).

LETTER

doi:10.1038/nature11291

A biophysical signature of network affiliation and sensory processing in mitral cells

Kamilla Angelo[1,2], Ede A. Rancz[1,3], Diogo Pimentel[1], Christian Hundahl[2,4], Jens Hannibal[4], Alexander Fleischmann[5], Bruno Pichler[3] & Troy W. Margrie[1,3]

One defining characteristic of the mammalian brain is its neuronal diversity[1]. For a given region, substructure, layer or even cell type, variability in neuronal morphology and connectivity persists[2–5]. Although it is well known that such cellular properties vary considerably according to neuronal type, the substantial biophysical diversity of neurons of the same morphological class is typically averaged out and ignored. Here we show that the amplitude of hyperpolarization-evoked sag of membrane potential recorded in olfactory bulb mitral cells is an emergent, homotypic property of local networks and sensory information processing. Simultaneous whole-cell recordings from pairs of cells show that the amount of hyperpolarization-evoked sag potential and current (I_h)[6] is stereotypic for mitral cells belonging to the same glomerular circuit. This is corroborated by a mosaic, glomerulus-based pattern of expression of the HCN2 (hyperpolarization-activated cyclic nucleotide-gated channel 2) subunit of the I_h channel. Furthermore, inter-glomerular differences in both membrane potential sag and HCN2 protein are diminished when sensory input to glomeruli is genetically and globally altered so that only one type of odorant receptor is universally expressed[7]. Population diversity in this intrinsic property therefore reflects differential expression between local mitral cell networks processing distinct odour-related information.

Neurons show a broad array of biophysical properties that profoundly impact the computations that they perform. Even within cell types, diversity in morphology[8], and expression of molecular markers[1] and ion channels[9] is well documented, but whether such variation reflects necessary biological noise[10] or perhaps a functional, dynamic system for regulating excitability at the cellular[11] or even network level remains unclear. The hyperpolarization-activated current (I_h; also observed as a sag potential) is one intrinsic biophysical property that is known to influence the input–output function[12–17] of most principal cell types[18]. In the olfactory bulb, the broad amplitude distribution of I_h-mediated sag potential recorded across the mitral cell population has recently been shown to reflect functional diversity in their input–output responses to *in vivo* stimuli[6].

To explore directly whether cell-to-cell variability in membrane potential sag might reflect differences between functional ensembles of mitral cells, we have taken advantage of the fact that in some brain regions the local architecture facilitates the identification of functionally discrete networks of neurons[19]. This is particularly true for the olfactory bulb, where glomeruli act as information hubs that receive inputs from a unique, homogeneous population of sensory afferents[20] that are integrated by a network of a few hundred interconnected local interneurons and principal mitral and tufted cells. Thus, in a slice preparation, individual mitral cells can be precisely linked to the functional circuit in which they operate *in vivo*[21,22], permitting us to explore whether their intrinsic diversity reflects an emergent property of the functional organization of the olfactory bulb (Fig. 1A).

In wild-type mice, mitral cells can exhibit a hyperpolarization-evoked rebound potential and current indicative of the I_h-mediated sag potential recently described in the rat, both *in vitro* and *in vivo*[6] (Supplementary Fig. 1). The distribution of sag potential amplitude (SPA) recorded across the mitral cell population in mouse is similarly uni-modal ($P < 0.05$) and broad (min = −10.4 mV, max = 30.65 mV, median = 3 mV, mean = 3.43 ± 5.8 mV, $n = 105$ cells, $n = 39$ animals; Fig. 1B and Supplementary Fig. 1). To explore the possibility that such population diversity might reflect differences between local mitral cell networks (Fig. 1A), we performed simultaneous whole-cell recordings of sag from pairs of cells belonging to distinctly different (Fig. 1C, a) or the same glomerular ensemble (Fig. 1C, b). The mean SPA found under these two recording scenarios was not significantly different (inter-glomerular pairs, 3.3 ± 6.7 mV, $n = 52$ cells versus intra-glomerular pairs, 2.45 ± 3.72 mV, $n = 28$ cells ($P = 0.41$); Fig. 1D). For each recorded pair we determined the absolute difference in SPA (Supplementary Fig. 2) and performed a multiple pair-wise comparison, whereby the SPA difference between each cell and all other cells within the same group—excluding its simultaneously recorded 'partner'—was calculated ('pseudo pairs', Fig. 1E). For inter-glomerular pairs of mitral cells, the distribution of SPA difference between recorded and pseudo pairs was similar (recorded: min = 0.03 mV, max = 21.06 mV, median = 3.57 mV, Q1 = 1.87 mV, Q3 = 5.35 mV, $n = 26$ pairs, versus pseudo pairs: min = 0.01 mV, max = 41.05 mV, median = 4.435 mV, Q1 = 1.78 mV, Q3 = 9.46 mV, $n = 1300$ ($P = 0.16$); Fig. 1E, F). This was also the case when comparing inter-glomerular recorded pairs and pseudo pairs extracted from our entire data set ($n = 105$ cells, 5,460 pseudo pairs; Supplementary Fig. 2).

In contrast, the sag potential and the I_h-current amplitude recorded simultaneously from mitral cells belonging to the same glomerular network were virtually indistinguishable (Supplementary Fig. 2). Thus, the difference in the SPA recorded from intra-glomerular pairs was significantly smaller than that determined for intra-glomerular pseudo pairs (recorded SPA difference for intra-glomerular pairs: min = 0 mV, max = 3.59 mV, median = 1.22 mV, Q1 = 0.31, Q3 = 2.1 mV, $n = 14$ pairs, versus intra-glomerular pseudo pairs: min = 0 mV, max = 12.75 mV, median = 4.23 mV, Q1 = 1.99, Q3 = 6.41 mV, $n = 364$ pairs ($P = 0.00002$); Fig. 1E, F and Supplementary Fig. 2), for inter-glomerular recorded pairs ($P = 0.024$; Fig. 1F, G) and pseudo pairs extracted from all cells ($n = 105$ cells; $P = 0.00001$; Fig. 1F and Supplementary Fig. 2). The broad range of sag and I_h-current amplitudes recorded across the bulb therefore reflects differences between individual glomerular circuits, in which it is a homotypic feature of the local mitral cell network (Fig. 1G).

Cell-attached experiments in mitral cell apical dendrites indicate that the I_h current is largest in patches recorded in the very distal region[6]. Thus, the recorded membrane potential sag may reflect activation of HCN channels that are expressed predominantly in the

[1]Department of Neuroscience, Physiology and Pharmacology, University College London, Gower Street, London WC1E 6BT, UK. [2]Department of Neuroscience and Pharmacology, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen N, Denmark. [3]Division of Neurophysiology, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. [4]Department of Clinical Biochemistry, Bispebjerg Hospital; University of Copenhagen, 2200 Copenhagen N, Denmark. [5]Center for Interdisciplinary Research in Biology (CIRB), College de France, 11 Place Marcelin Berthelot, 75231 Paris Cedex 05, France.

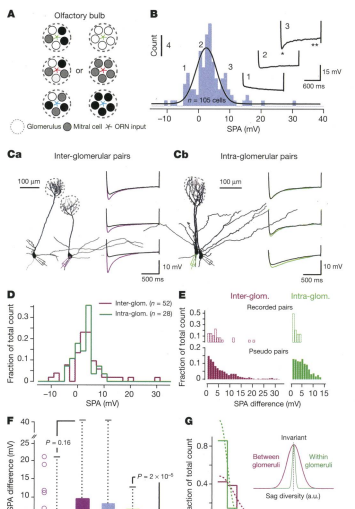


Figure 1 | Diversity of sag potential amplitude within and between mitral cell networks. **A**, Schematic of two possible scenarios underlying the population diversity of mitral cell sag. Each olfactory glomerulus receives genetically unique input (shown in red, green and blue). Membrane potential sag (shown by different coloured circles) may be hetero- (left) or homogeneously (right) expressed with the glomerulus. ORN, olfactory receptor nerve. **B**, Histogram (fitted with a Gaussian curve) showing the distribution of SPA recorded across the mitral cell population ($n = 105$ cells). Traces, three examples recorded from cells belonging to the indicated bins. The sag potential amplitude is determined as the voltage difference between the peak hyperpolarization (shown by an asterisk) and the steady-state membrane potential (shown by two asterisks). **Ca**, **b**, Example morphologies of two simultaneously recorded mitral cells projecting to either different (**Ca**) or the same (**Cb**) glomerular networks. The voltage traces show the sag potential recorded in three examples of inter- and intra-glomerular pairs. **D**, Histograms of the SPA for all individual cells belonging to either inter- and intra-glomerular pairs. **E**, Top panel, histograms of recorded SPA differences for inter- ($n = 26$) and intra-glomerular ($n = 14$) pairs. Bottom panel, histograms of SPA differences for inter- and intra-glomerular pseudo pairs. **F**, Box-plot of recorded and pseudo inter-glomerular pairs, pseudo pairs of all recordings, and intra-glomerular pseudo and recorded pairs. Open circles, individual data points of recorded SPA difference. **G**, Histogram of sag amplitude difference for inter- and intra-glomerular recorded pairs fitted with a half Gaussian ($n = 26$ and 14 pairs, respectively; bin size, 3 mV). Inset, Gaussian fits of the recorded data for intra-glomerular and inter-glomerular pairs. a.u., arbitrary units.

dendritic tuft, the site of sensory integration within the glomerulus. To explore this possibility, we performed immunohistochemical staining for the HCN2 subunit that can form both homomeric or heteromeric HCN channels known to mediate the slow I_h (ref. 23) underlying mitral cell sag[6]. Qualitatively, very little HCN2 protein was seen in the granule cell and mitral cell layers[24]. In contrast to the low but homogeneous HCN2 expression in the external plexiform layer[24], we found a high-contrast mosaic staining pattern across the glomerular layer (Fig. 2a). To determine whether HCN2 expression within the glomerulus was postsynaptic to olfactory receptor neuron input we next used a transgenic mouse line that expresses the tau-LacZ transgene in the sensory afferents under the olfactory marker protein (OMP) promoter (OMP-IRES-tau-LacZ mice)[7]. Double-staining experiments against both LacZ and HCN2 showed that the HCN2 protein is predominantly expressed in dendritic compartments within the glomerulus and downstream of the olfactory receptor input (Fig. 2b–d)[25]. Irrespective of the potential contribution of other cell types[24,26], this mosaic pattern of HCN2 expression is consistent with the observation of large I_h currents in the distal apical dendrite and the broad range of sag amplitudes recorded in mitral cells participating in different glomerular networks.

Such glomerular-based SPA and HCN2 expression might reflect network-related homeostatic regulation[27,28] of excitability, in which glomerular differences arise from the processing of functionally and

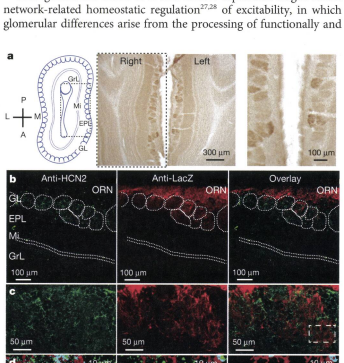


Figure 2 | Glomerular expression of HCN2 in wild-type and OMP-IRES-tau-LacZ mice. **a**, Left panel, schematic of a horizontal section of an olfactory bulb highlighting its anatomical organization. A, anterior; L, lateral; M, medial; P, posterior. Middle and right panels, HCN2–DAB staining highlighting glomerular diversity. **b**, Left and middle panels, anti-HCN2 and anti-β-galactosidase staining in OMP-IRES-tau-LacZ mice. Right panel, green and red channels shown together for comparison. **c**, High-magnification images of the glomerular layer (colours as in **b**). **d**, Left, higher-magnification images scanned as 30 digital sections, shown as a pseudo three-dimensional image of an approximately 6-µm thick section. Areas of overlap between green and red are shown in light blue and indicate close proximity of the HCN2 and OMP-LacZ signals. Middle and right panels, images of the outlined area shown in the left panel (anti-HCN2 staining (green colour) has been removed for clarity). EPL, external-plexiform layer; GL, glomerular layer; GrL, granule-cell layer; Mi, mitral cell layer; ORN, olfactory receptor nerve layer.

genetically unique subsets of olfactory input[20]. To test this hypothesis we next performed experiments in a transgenic 'monoclonal nose' mouse that expresses the M71 odorant receptor in more than 95% of receptor neurons[7] (M71tg mice, Fig. 3A). We calculated mean pixel intensities of glomeruli in HCN2–DAB-stained M71tg and control animals that revealed significantly different variances ($P = 0.001$) whereby the mean pixel intensity for glomeruli varied less in the M71tg mice than in the control mice ($P = 0.0025$; Fig. 3A, Supplementary Fig. 3). Simultaneous recordings from inter-glomerular pairs of mitral cells in M71tg mice (min = 0.08 mV, max = 6.3 mV, median = 3.3 mV, Q1 = 1.32 mV, Q3 = 3.78 mV; $n = 24$; Fig. 3B, a) also revealed a significantly narrower distribution of SPA difference compared to wild-type and control mice (min = 0.23 mV, max = 21.04 mV, median = 3.4 mV, Q1 = 1.78 mV, Q3 = 10.78 mV, $n = 24$ ($P = 0.004$); Fig. 3B, a and b, C). This reduction in sag amplitude diversity was also notable for pseudo-pair comparisons for all recorded pairs (Fig. 3D, a) and the overall population data set (M71tg mice: min = 0 mV, max = 36.01 mV, median = 3.4 mV, Q1 = 1.6, Q3 = 6.1 mV, $n = 91$ cells, 4,095 comparisons, versus M71 control mice: min = 0 mV, max = 51.4 mV, median = 4.6 mV, Q1 = 2, Q3 = 10.19 mV, $n = 81$ cells, 3,321 comparisons ($P < 4 \times 10^{-47}$); Fig. 3D, a and c). Thus, mitral cells and glomeruli in M71tg mice are more homogeneous in their sag and HCN2 expression profile than those in wild-type and M71 control mice receiving the normal, genetically diverse, array of olfactory receptor neuron input.

Despite the overall reduction in SPA variance in the M71tg mice, the SPA in the intra-glomerular pairs remained more similar (M71tg intra-glomerular pairs: min = 0.2 mV, max = 6.5 mV, median = 0.8 mV, Q1 = 0.3 mV, Q3 = 1.26 mV; $n = 9$, versus M71tg inter-glomerular pairs ($P = 0.04$); Figs 3C and 4a). Indeed, we observed no effect of wholesale expression of the M71 receptor on intra-glomerular sag diversity (M71tg versus control and wild-type mice; $P = 0.61$ and 0.46, respectively; Fig. 4a). Thus, sensory afferent input seems unlikely to be the sole driver of inter-glomerular diversity (Fig. 4b).

Using the hyperpolarization-evoked sag potential as a general proxy, we have identified several organizing principles regarding the population diversity of I_h expression in mitral cells[6,29]. First, this intrinsic property is a biophysical fingerprint of local constellations of mitral cells forming a functionally discrete olfactory network (Fig. 4c). As mitral cells are electrically and exclusively coupled to their intra-glomerular counterparts, co-regulation of the I_h channel and current via gap junctions may contribute to their biophysical similarity[6]. Second, analysis of the M71tg mouse shows directly that this network-affiliation-based signature depends on sensory information processing. The fact that in the M71tg mouse intra-glomerular sag diversity remained more homogeneous than in the overall population also suggests that other factors such as feed-forward and lateral inhibition may contribute to sag regulation at the level of the glomerulus. From a functional perspective, mitral cell F–I curves (frequency of spiking output of neurons (F) in response to varying current injection

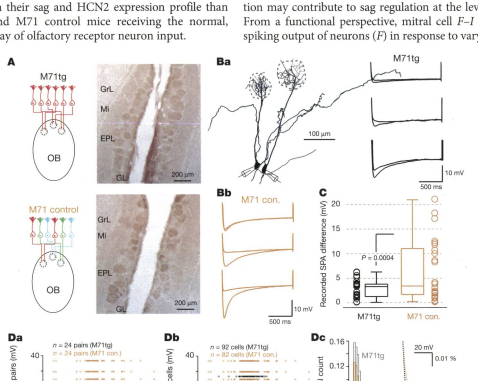


Figure 3 | Glomerular expression of HCN2 and mitral cell sag in M71 monoclonal mice. **A**, Schematics of the genetic organization of the nose to brain connectivity in M71tg and control mice (left panels). In contrast to the normal genetically diverse ORNs and glomerulus-specific projection pattern in wild-type and control animals, the sensory afferents to all glomeruli in the transgenic mice express the same M71 receptor (red). HCN2–DAB staining (right panels) in horizontal olfactory bulb (OB) slices from an M71tg mouse and a M71 control mouse. **Ba**, **b**, Example morphologies of two simultaneously recorded mitral cells belonging to distinctly different glomeruli (inter-glomerular pair) in the M71 transgenic mouse. The voltage traces show the sag potential recorded in three different example pairs in M71tg (**Ba**) and control mice (**Bb**). **C**, Box-plot of the SPA difference for inter-glomerular recorded pairs from M71tg and control mice. Open circles, individual data points of recorded SPA difference. **Da**, **b**, Scatter-plots of SPA for pseudo pairs extracted from paired inter-glomerular recordings in the M71tg and control mice. **Dc**, Histogram of SPA differences for pseudo pairs extracted from all cells recorded in the M71tg ($n = 91$ cells, 4,095 comparisons) and control mice ($n = 81$ cells, 3,321 comparisons; bin size, 1 mV). Inset, histogram spline fits are overlaid and zoomed in to highlight the disparity between the distributions.

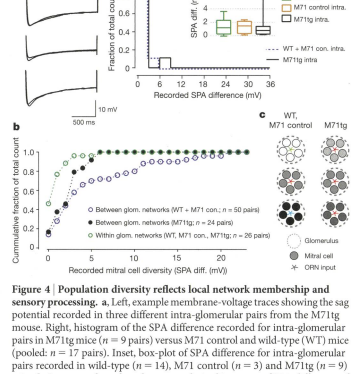


Figure 4 | Population diversity reflects local network membership and sensory processing. a, Left, example membrane-voltage traces showing the sag potential recorded in three different intra-glomerular pairs from the M71tg mouse. Right, histogram of the SPA difference recorded for intra-glomerular pairs in M71tg mice ($n = 9$ pairs) versus M71 control and wild-type (WT) mice (pooled: $n = 17$ pairs). Inset, box-plot of SPA difference for intra-glomerular pairs recorded in wild-type ($n = 14$), M71 control ($n = 3$) and M71tg ($n = 9$) mice. **b**, Summary data plotted as a cumulative histogram of SPA difference of all recorded pairs. **c**, Schematic highlighting the relationship between glomerulus affiliation, sensory input and mitral cell sag diversity.

amplitudes (D)) are known to shift left, from a sigmoid towards a linear operation with increasing sag[6]. We suggest that the glomerular basis of this delineation may therefore reflect a network-based gain control mechanism and contribute to correlated output patterning at the level of mitral cell networks[30]. Irrespective of the cellular mechanisms underlying this glomerular phenomenon, the network-based regulation of this mitral cell property appears fundamental to the organization and function of olfactory bulb circuits.

METHODS SUMMARY

Whole-cell recordings using standard intracellular and extracellular solution were carried out in horizontal olfactory bulb slices (300-μm thick) prepared from wild-type C57Bl/6J and M71 transgenic or control littermate mice aged 4–6 weeks. The I_h current and I_h-mediated sag potential has recently been extensively characterized for mitral cells[6], and the details of the experimental and analytical procedures are provided in the Supplementary Information.

Received 7 August 2011; accepted 7 June 2012.
Published online 22 July 2012.

1. Gupta, A., Wang, Y. & Markram, H. Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science* **287,** 273–278 (2000).
2. Brochtrup, A. & Hummel, T. Olfactory map formation in the *Drosophila* brain: genetic specificity and neuronal variability. *Curr. Opin. Neurobiol.* **21,** 85–92 (2011).
3. Reyes, A. *et al.* Target-cell specific facilitation and depression in neocortical networks. *Nature Neurosci.* **1,** 279–285 (1998).
4. Jinno, S. *et al.* Neuronal diversity in GABAergic long-range projections from the hippocampus. *J. Neurosci.* **27,** 8790–8804 (2007).
5. Brown, S. P. & Hestrin, S. Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* **457,** 1133–1136 (2009).
6. Angelo, K. & Margrie, T. W. Population diversity and function of hyperpolarization-activated current in olfactory bulb mitral cells. *Scientific Reports* **1,** 1:50 (2011).
7. Fleischmann, A. *et al.* Mice with a "monoclonal nose": perturbations in an olfactory map impair odor discrimination. *Neuron* **60,** 1068–1081 (2008).
8. Tsiola, A., Hamzei-Sichani, F., Peterlin, Z. & Yuste, R. Quantitative morphologic classification of layer 5 neurons from mouse primary visual cortex. *J. Comp. Neurol.* **461,** 415–428 (2003).
9. Schulz, D. J., Goaillard, J. M. & Marder, E. Variable channel expression in identified single and electrically coupled neurons in different animals. *Nature Neurosci.* **9,** 356–362 (2006).
10. Ermentrout, G. B., Galan, R. F. & Urban, N. N. Reliability, synchrony and noise. *Trends Neurosci.* **31,** 428–434 (2008).
11. Marder, E. & Goaillard, J. M. Variability, compensation and homeostasis in neuron and network function. *Nature Rev. Neurosci.* **7,** 563–574 (2006).
12. Magee, J. C. Dendritic hyperpolarization-activated currents modify the integrative properties of hippocampal CA1 pyramidal neurons. *J. Neurosci.* **18,** 7613–7624 (1998).
13. Garden, D. L., Dodson, P. D., O'Donnell, C., White, M. D. & Nolan, M. F. Tuning of synaptic integration in the medial entorhinal cortex to the organization of grid cell firing fields. *Neuron* **60,** 875–889 (2008).
14. George, M. S., Abbott, L. F. & Siegelbaum, S. A. HCN hyperpolarization-activated cation channels inhibit EPSPs by interactions with M-type K^+ channels. *Nature Neurosci.* **12,** 577–584 (2009).
15. Nolan, M. F., Dudman, J. T., Dodson, P. D. & Santoro, B. HCN1 channels control resting and active integrative properties of stellate cells from layer II of the entorhinal cortex. *J. Neurosci.* **27,** 12440–12451 (2007).
16. Lüthi, A. & McCormick, D. A. Periodicity of thalamic synchronized oscillations: the role of Ca^{2+}-mediated upregulation of I_h. *Neuron* **20,** 553–563 (1998).
17. Giocomo, L. M., Zilli, E. A., Fransen, E. & Hasselmo, M. E. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science* **315,** 1719–1722 (2007).
18. Migliore, M. & Shepherd, G. M. Emerging rules for the distributions of active dendritic conductances. *Nature Rev. Neurosci.* **3,** 362–370 (2002).
19. Woolsey, T. A. & Van der Loos, H. The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex. The description of a cortical field composed of discrete cytoarchitectonic units. *Brain Res.* **17,** 205–242 (1970).
20. Mombaerts, P. Targeting olfaction. *Curr. Opin. Neurobiol.* **6,** 481–486 (1996).
21. Schoppa, N. E. & Westbrook, G. L. Glomerulus-specific synchronization of mitral cells in the olfactory bulb. *Neuron* **31,** 639–651 (2001).
22. Pimentel, D. O. & Margrie, T. W. Glutamatergic transmission and plasticity between olfactory bulb mitral cells. *J. Physiol.* **586,** 2107–2119 (2008).
23. Robinson, R. B. & Siegelbaum, S. A. Hyperpolarization-activated cation currents: from molecules to physiological function. *Annu. Rev. Physiol.* **65,** 453–480 (2003).
24. Notomi, T. & Shigemoto, R. Immunohistochemical localization of Ih channel subunits, HCN1–4, in the rat brain. *J. Comp. Neurol.* **471,** 241–276 (2004).
25. Santoro, B. *et al.* Molecular and functional heterogeneity of hyperpolarization-activated pacemaker channels in the mouse CNS. *J. Neurosci.* **20,** 5264–5275 (2000).
26. Nicoll, L. & BellUzzi, O. Hyperpolarisation-activated current in glomerular cells of the rat olfactory bulb. *Neuroreport* **12,** 3117–3120 (2001).
27. van Welie, I., van Hooft, J. A. & Wadman, W. J. Homeostatic scaling of neuronal excitability by synaptic modulation of somatic hyperpolarization-activated Ih channels. *Proc. Natl Acad. Sci. USA* **101,** 5123–5128 (2004).
28. Desai, N. S., Rutherford, L. C. & Turrigiano, G. G. Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature Neurosci.* **2,** 515–520 (1999).
29. Padmanabhan, K. & Urban, N. N. Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nature Neurosci.* **13,** 1276–1282 (2010).
30. Dhawale, A. K., Hagiwara, A., Bhalla, U. S., Murthy, V. N. & Albeanu, D. F. Non-redundant odor coding by sister mitral cells revealed by light addressable glomeruli in the mouse. *Nature Neurosci.* **13,** 1404–1412 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Velez-Fort for comments on the manuscript. E.A.R. is a recipient of a Sir Henry Wellcome Fellowship. This project was supported by the Oticon Foundation, the Danish Council for Independent Research and the Lundbeckfondation (K.A.), the Gulbenkian PhD Programme and Fundação para a Ciência e Tecnologia (D.P.), The Wellcome Trust (T.W.M.) and Medical Research Council MC_U1175975156 (B.P. and T.W.M.).

Author Contributions K.A. and E.A.R. performed electrophysiological experiments. C.H. and J.H. carried out immunohistochemistry. D.P. performed morphological reconstructions and some initial electrophysiological experiments. B.P. and E.A.R. contributed to data analysis. A.F. generated the transgenic mouse line. K.A. and T.W.M. conceived the project and performed analysis. T.W.M. wrote the paper with input from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.W.M. (troy.margrie@nimr.mrc.ac.uk).

LETTER

doi:10.1038/nature11312

Activation of specific interneurons improves V1 feature selectivity and visual perception

Seung-Hee Lee[1,2], Alex C. Kwan[1], Siyu Zhang[1], Victoria Phoumthipphavong[1], John G. Flannery[3], Sotiris C. Masmanidis[3], Hiroki Taniguchi[4], Z. Josh Huang[4], Feng Zhang[5], Edward S. Boyden[6], Karl Deisseroth[5] & Yang Dan[1,2]

Inhibitory interneurons are essential components of the neural circuits underlying various brain functions. In the neocortex, a large diversity of GABA (γ-aminobutyric acid) interneurons has been identified on the basis of their morphology, molecular markers, biophysical properties and innervation pattern[1–3]. However, how the activity of each subtype of interneurons contributes to sensory processing remains unclear. Here we show that optogenetic activation of parvalbumin-positive (PV^+) interneurons in the mouse primary visual cortex (V1) sharpens neuronal feature selectivity and improves perceptual discrimination. Using multichannel recording with silicon probes[4,5] and channelrhodopsin-2 (ChR2)-mediated optical activation[6], we found that increased spiking of PV^+ interneurons markedly sharpened orientation tuning and enhanced direction selectivity of nearby neurons. These effects were caused by the activation of inhibitory neurons rather than a decreased spiking of excitatory neurons, as archaerhodopsin-3 (Arch)-mediated optical silencing[7] of calcium/calmodulin-dependent protein kinase IIα (CAMKIIα)-positive excitatory neurons caused no significant change in V1 stimulus selectivity. Moreover, the improved selectivity specifically required PV^+ neuron activation, as activating somatostatin or vasointestinal peptide interneurons had no significant effect. Notably, PV^+ neuron activation in awake mice caused a significant improvement in their orientation discrimination, mirroring the sharpened V1 orientation tuning. Together, these results provide the first demonstration that visual coding and perception can be improved by increased spiking of a specific subtype of cortical inhibitory interneurons.

To allow specific activation of PV^+ interneurons, we injected a Cre-inducible adeno-associated virus (AAV) vector containing the ChR2 gene fused in-frame with the coding sequence for enhanced yellow fluorescent protein (eYFP; Supplementary Fig. 1a)[8] into the V1 of a knock-in mouse line expressing Cre recombinase under the PV promoter (PV-ChR2; see Methods)[9,10]. Three weeks after injection, immunostaining confirmed eYFP expression specific to PV^+ neurons (Fig. 1a). To measure the effect of ChR2 activation, we inserted a multichannel silicon probe[4,5] near to the injection site for simultaneous recording from all cortical layers (Supplementary Fig. 1b). After stimulation with blue (473 nm) laser, a small fraction (12 out of 96; 13%) of the neurons showed significant increases in spontaneous firing rate, whereas 43% (41 out of 96) showed significant decreases (Fig. 1b, c and

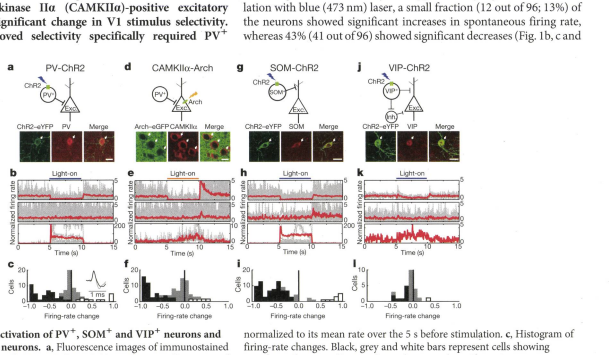


Figure 1 | Optogenetic activation of PV^+, SOM^+ and VIP^+ neurons and silencing of $CAMKII\alpha^+$ neurons. a, Fluorescence images of immunostained PV^+ cells (red) expressing ChR2–eYFP (green). Top, schematic illustration of experiment in **a**–**c**. **b**, Peristimulus time histograms of neurons during 30 repeats of blue-laser stimulation. Top, cells showing a significant firing-rate decrease ($P < 0.01$, bootstrap; $n = 41$); middle, cells without significant change ($n = 43$); bottom, cells with a significant increase ($n = 12$). Grey denotes individual cells and red denotes the average within each group. Blue bar indicates the duration of laser stimulation (5 s). The firing rate of each cell was normalized to its mean rate over the 5 s before stimulation. **c**, Histogram of firing-rate changes. Black, grey and white bars represent cells showing significant decreases ($P < 0.01$), no change and significant increases, respectively. Inset, spike waveform averaged across cells with significantly decreased (black) or increased (grey) firing. **d**–**f**, Similar to **a**–**c** for Arch-mediated silencing of $CAMKII\alpha^+$ neurons. **g**–**i**, Similar to **a**–**c** for ChR2-mediated activation of SOM^+ neurons. **j**–**l**, Similar to **a**–**c** for ChR2-mediated activation of VIP^+ neurons. Exc., excitatory neuron; Inh., inhibitory neuron. Thunderbolt marks denote laser stimulation. Scale bars, 15 μm.

[1]Division of Neurobiology, Department of Molecular and Cell Biology, Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA. [2]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. [3]Broad Fellows Program in Brain Circuitry, Kavli Nanoscience Institute, California Institute of Technology, Pasadena, California 91125, USA. [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [5]Department of Bioengineering, Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA. [6]Media Laboratory, Biological Engineering, Brain and Cognitive Sciences, and McGovern Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Supplementary Fig. 1c). The cells with increased firing showed narrower spike waveforms than those with decreased firing (Fig. 1c, inset; peak–trough width, 0.31 ± 0.12 versus 0.44 ± 0.09 ms; mean $\pm$ s.d.), indicating that the PV^+ interneurons activated directly by laser were mostly fast-spiking cells[11]. The decreased spiking of other neurons is probably caused by increased inhibition from the activated PV^+ neurons.

We then measured orientation tuning and direction selectivity of the neurons using drifting grating stimuli, both with and without ChR2 activation. We found that PV^+ neuron activation caused considerable changes in the tuning of other neurons. In addition to the overall firing-rate decrease, we also observed a marked reduction in orientation tuning width and an increase in direction selectivity (Fig. 2a). To quantify these effects, we fitted each tuning curve with a double-Gaussian function (Supplementary Fig. 2a). For all the well-tuned neurons (the tuning curves of which with and without ChR2 activation were well fitted; see Methods), ChR2 activation caused a significant decrease in tuning bandwidth (σ; $P < 10^{-4}$, paired t-test; Fig. 2b) and an increase in direction selectivity index (DSI; $P = 0.007$; Fig. 2c). However, the preferred orientation (θ_0) remained relatively constant (Fig. 2d). Interestingly, all of the putative PV^+ neurons (Fig. 1c, white bars) showed poor tuning even before laser stimulation, consistent with previous findings[12,13]. Although most of the poorly fitted tuning curves showed low signal-to-noise ratio (Supplementary Fig. 2b, c, examples 1, 2), a few had relatively reliable tuning curves showing several irregular peaks (Supplementary Fig. 2b, c, examples 3–6). Laser stimulation also sharpened the tuning of some of these cells, although the degree of sharpening was not well quantified by curve fitting.

The decrease in σ was correlated with the reduction in firing rate (Supplementary Fig. 3e), suggesting that the enhanced selectivity depends on the increase in inhibition. However, we also wondered whether an overall reduction in cortical firing is sufficient to enhance the selectivity without interneuron activation. As an alternative method to decrease cortical spiking, we expressed the light-activated proton pump Arch$^+$ in $CAMKII\alpha^+$ excitatory neurons by injecting AAV (Supplementary Fig. 1a) into the V1 of the CAMKIIα-Cre mice[14] (CAMKIIα-Arch; Fig. 1d). Arch-mediated partial silencing of $CAMKII\alpha^+$ neurons caused decreases in the spontaneous (Fig. 1e, f and Supplementary Fig. 1c) and visually evoked (Supplementary Fig. 1c, d) firing rates, similar to PV^+ neuron activation. However, $CAMKII\alpha^+$ neuron silencing caused no significant change in σ ($P = 0.75$; Fig. 2f) or DSI ($P = 0.68$; Fig. 2g). In addition, there was no significant correlation between the firing-rate reduction and change in σ or DSI (Supplementary Fig. 3f, j).

Next, we wanted to test whether the enhanced selectivity required a general increase in inhibition or specific activation of PV^+ interneurons.

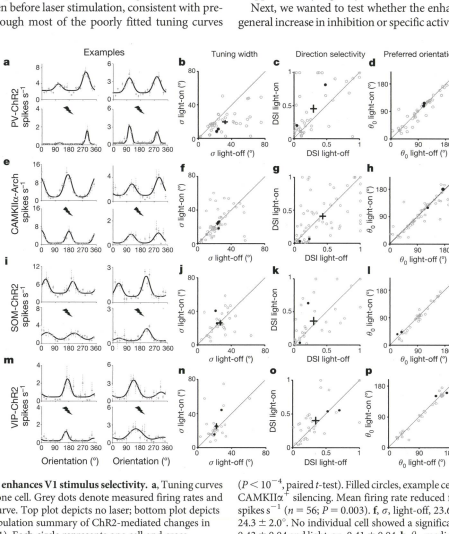


Figure 2 | PV^+ activation enhances V1 stimulus selectivity. **a**, Tuning curves of two cells, each box for one cell. Grey dots denote measured firing rates and black line denotes fitted curve. Top plot depicts no laser; bottom plot depicts laser stimulation. **b**–**d**, Population summary of ChR2-mediated changes in stimulus selectivity ($n = 41$). Each circle represents one cell and cross represents population average. **b**, Tuning width (σ), light-off, $32.1 \pm 2.9^\circ$ and light-on, $19.7 \pm 1.9^\circ$; 22% individual neurons showed a significant decrease at $P < 0.01$ (bootstrap) and none showed a significant increase. **c**, DSI, light-off, 0.30 ± 0.05 and light-on, 0.45 ± 0.06; 11% of neurons showed a significant increase and none showed a significant decrease. **d**, Preferred orientation (θ_0), median difference between light-on and light-off, 6.2°. For neurons in **b**–**d**, the laser reduced visually evoked firing rate from 5.5 ± 0.6 to 2.8 ± 0.5 spikes s^{-1} ($P < 10^{-4}$, paired t-test). Filled circles, example cells in **a**. **e**–**h**, Similar to **a**–**d** for $CAMKII\alpha^+$ silencing. Mean firing rate reduced from 3.0 ± 0.4 to 2.2 ± 0.4 spikes s^{-1} ($n = 56$; $P = 0.003$). **f**, σ, light-off, $23.6 \pm 1.8^\circ$ and light-on, $24.3 \pm 2.0^\circ$. No individual cell showed a significant change. **g**, DSI, light-off, 0.43 ± 0.04 and light-on, 0.41 ± 0.04. **h**, θ_0, median difference, 5.4°. **i**–**l**, Similar to **a**–**d** for SOM-ChR2 mice. Firing rate reduced from 3.4 ± 0.6 to 1.9 ± 0.4 spikes s^{-1} ($P < 10^{-3}$). **j**, σ, light-off, $26.9 \pm 2.1^\circ$ and light-on, $26.2 \pm 2.3^\circ$, $n = 33$. No individual cell showed a significant change. **k**, DSI, light-off, 0.29 ± 0.05 and light-on, 0.36 ± 0.06. **l**, θ_0, median difference, 5.7°. **m**–**p**, Similar to **a**–**d** for VIP-ChR2 mice ($n = 31$). **n**, σ, light-off, $21.8 \pm 1.8^\circ$ and light-on, $25.5 \pm 2.8^\circ$. **o**, DSI, light-off, 0.35 ± 0.05 and light-on, 0.41 ± 0.05. **p**, θ_0, median difference, 5.6°. All data are mean $\pm$ s.e.m.

In addition to PV^+ neurons, somatostatin $(SOM)^+$ and vasointestinal peptide $(VIP)^+$ neurons constitute two other major subtypes of GABAergic interneurons[2]. To test their roles in shaping V1 selectivity, we induced cell-type-specific expression of ChR2–eYFP in SOM-Cre or VIP-Cre mice (Fig. 1g, j). In the SOM-ChR2 mice, the laser also caused increased spiking of a few neurons (11 out of 91) but decreased firing in most of the neurons (71 out of 91) (Fig. 1h, i, and Supplementary Fig. 1c, d). In fact, compared with the PV-ChR2 group, a higher percentage of neurons showed decreased firing. However, SOM^+ activation caused no significant effect on either σ ($P = 0.79$; Fig. 2i, j) or DSI ($P = 0.17$; Fig. 2i, k). In VIP-ChR2 mice, laser stimulation also induced no significant change in σ ($P = 0.16$; Fig. 2m, n) or DSI ($P = 0.16$; Fig. 2m, o), and only a mild change in firing rate (Fig. 1k, l and Supplementary Fig. 1c, d). The change in σ showed a negative correlation with firing-rate change (Supplementary Fig. 3h), opposite to PV^+ activation (Supplementary Fig. 3e). This may be because the VIP^+ neurons also innervate inhibitory interneurons[15], thus causing both inhibition and disinhibition. In addition to σ, tuning measured by orientation selectivity index was also improved by PV^+ activation, but not by $CAMKII\alpha^+$ silencing or SOM^+ or VIP^+ activation (Supplementary Fig. 4).

Even among the PV-ChR2, SOM-ChR2 and CAMKIIα-Arch groups, the degrees of firing-rate suppression were not identical. To ensure that the difference in tuning width reduction was not caused by differences in firing-rate suppression, we selected neurons from each group that showed intermediate levels of suppression (firing-rate change, -0.75 to -0.3; Supplementary Fig. 5a). Although the spike-rate suppression is well matched across groups within this range (mean rate change, -0.51 (PV), -0.50 (SOM), -0.50 (CAMKIIα)), laser stimulation caused a significant decrease in σ in the PV (-0.19 ± 0.05; mean $\pm$ s.e.m.) but not the SOM (0.01 ± 0.09) or CAMKIIα (-0.05 ± 0.03) groups. The magnitudes of reduction in σ were significantly different between the PV and SOM or CAMKIIα groups (Supplementary Fig. 5b). We also noticed that neurons in these groups showed different initial tuning widths ($\sigma_{light-off}$). To ensure that the difference in tuning width reduction was not caused by the difference in $\sigma_{light-off}$, we selected neurons with $\sigma_{light-off}$ falling within two ranges: the broad (25–50°) and narrow (0–25°) ranges. Within each range, the median values of $\sigma_{light-off}$ were not significantly different across groups, but laser stimulation caused a significant decrease in σ in the PV-ChR2 group (broad, -0.28 ± 0.09; narrow, -0.11 ± 0.05) and not in any other group (Supplementary Fig. 5c, d). Furthermore, within each range the amount of decrease in σ was significantly different between the PV group and other groups.

Although the lack of effect of VIP^+ activation on σ is not surprising given its moderate effect on firing rate, what underlies the difference between SOM^+ and PV^+ neurons is not obvious. Tuning of a sensory neuron depends on its firing rate versus input current (F–I) curve: subtractive modulation of the F–I curve causes sharpening of tuning, whereas divisive modulation has no effect[16]. We therefore tested the effects of SOM^+ and PV^+ activation on neuronal F–I functions using *in vivo* whole-cell recording (see Methods). For each neuron we measured firing rate versus injected depolarizing current. We found that PV^+ activation caused a large increase in spiking threshold ($P = 6 \times 10^{-4}$, paired t-test) without significantly changing the slope ($P = 0.34$; Fig. 3a–d), but SOM^+ activation caused a marked reduction of the slope ($P = 0.02$) without a significant change in threshold ($P = 0.18$; Fig. 3e–h). These effects may be related to the different subcellular targeting of PV^+ and SOM^+ neurons onto pyramidal neurons (perisomatic versus dendritic)[17]. When we plotted the median response of each group at each orientation with versus without laser stimulation, we also found a strong subtractive effect of PV^+ activation but a primarily divisive effect of SOM^+ activation (Supplementary Fig. 6).

Changes in V1 orientation tuning can, in principle, affect perceptual discrimination[18,19]. To test the functional consequence of sharpened tuning, we applied laser stimulation to awake PV-ChR2 mice while they performed orientation discrimination. The head-restrained mice were trained on a go/no-go task and rewarded for licking in response to a grating at one of two orientations[20] (see Methods, Fig. 4a, b and Supplementary Movie 1). For an easy discrimination task (difference between go and no-go orientations $\Delta\theta = 90°$), the performance measured by the divergence between hit and false-alarm rates (Fig. 4c) and discriminability (d'; Fig. 4d) improved systematically over days. As expected, d' increased monotonically with $\Delta\theta$ (Fig. 4e). After the performance reached a steady state, we tested the effect of laser and found a significant improvement in d' at a range of $\Delta\theta$ (Fig. 4f and Supplementary Fig. 7a–c). To ensure that the improvement was caused by optical activation of PV^+ neurons rather than through the retinal photoreceptors activated by scattered laser, we performed a control experiment in mice that were not virus-infected and found no significant increase in d' (Fig. 4g). Separate electrophysiological experiments in awake mice showed that PV^+ activation also caused a significant sharpening of tuning ($P < 0.05$; Supplementary Fig. 8), similar to that in anaesthetized mice (Fig. 2a, b). Thus, the sharpening of V1 orientation tuning induced by PV^+ activation can facilitate orientation discrimination. Unlike the learning curve shown in Fig. 4d, in which d' improved steadily over 5–10 days of training, the effect of PV^+ activation showed no systematic change over several

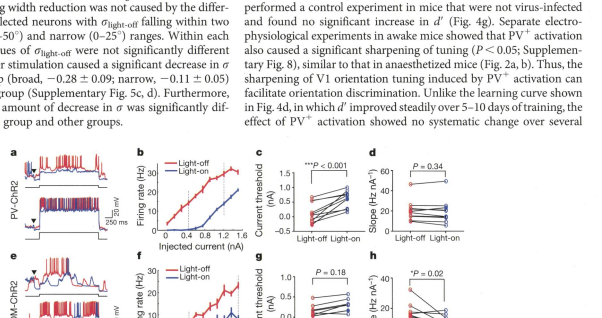


Figure 3 | Effects of PV^+ and SOM^+ activation on F–I function. **a**, Example traces showing neuronal spiking evoked by current injection. Red indicates voltage trace without laser and blue indicates voltage trace with laser. Arrowhead indicates laser onset (laser offset, 800 ms after current injection; not shown). Current amplitude, 0.45 nA (upper) and 1.2 nA (lower). **b**, F–I curve of the cell shown in **a**. Red denotes firing rate without laser and blue denotes firing rate with laser. Dashed lines indicate current amplitudes shown in **a**. **c**, **d**, Summary of threshold (lowest current that evokes spiking) and F–I slope with and without laser. Each symbol represents one cell ($n = 10$). **e**–**h**, Similar to **a**–**d** for SOM^+ activation ($n = 8$). All data are $\pm$ s.e.m.

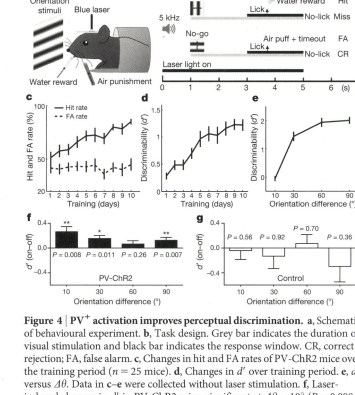


Figure 4 | PV^+ activation improves perceptual discrimination. a, Schematic of behavioural experiment. **b**, Task design. Grey bar indicates the duration of visual stimulation and black bar indicates the response window. CR, correct rejection; FA, false alarm. **c**, Changes in hit and FA rates of PV-ChR2 mice over the training period ($n = 25$ mice). **d**, Changes in d' over training period. **e**, d' versus $\Delta\theta$. Data in **c**–**e** were collected without laser stimulation. **f**, Laser-induced changes in d' in PV-ChR2 mice, significant at $\Delta\theta = 10°$ ($P = 0.008$, Wilcoxon signed-rank test, 0.032 after Bonferroni correction; $n = 9$ mice), $\Delta\theta = 30°$ ($P = 0.011$; $n = 19$) and $\Delta\theta = 90°$ ($P = 0.007$; $n = 25$). **g**, Similar to **f** but for control mice ($n = 10$). All data are mean ± s.e.m.

days of testing (Supplementary Fig. 7d), suggesting that the effect was not caused by learning of a distorted perception owing to PV^+ activation.

Previous pharmacological experiments have shown that the endogenous level of GABAergic inhibition is necessary for orientation tuning and direction selectivity[21,22]. Our study showed that increasing inhibitory neuron activity above the normal level can further sharpen cortical feature selectivity and improve perceptual discrimination. A recent study showed that PV^+ activation only moderately affects the tuning of V1 neurons[23]. This is probably caused by the relatively low level of PV^+ activation compared with our study, as the increase in stimulus selectivity is strongly correlated with the degree of firing-rate suppression (Supplementary Fig. 3e, i). Interestingly, studies in awake, behaving primates have shown that increases in task difficulty[24] and attention[25] are accompanied by a preferential increase in the firing of fast-spiking neurons, many of which are likely to be PV^+ interneurons[11]. This suggests that the level of inhibitory activity can be dynamically regulated by top-down mechanisms to meet changing behavioural demands.

A long-standing debate on the mechanism for orientation tuning is whether the excitatory and inhibitory inputs show similar tuning[26]. In the rodent V1, whether inhibitory interneurons are well tuned has also been controversial[12,13,27]. Our results show that an increase in the untuned activity of inhibitory interneurons is sufficient to sharpen orientation tuning, consistent with previous theoretical prediction[28]. Furthermore, although driving either PV^+ or SOM^+ interneurons markedly suppressed cortical spiking, only PV^+ activation could sharpen orientation tuning. This specificity may be attributable to the powerful perisomatic inhibition exerted by PV^+ neurons[1,29]. Recent studies have demonstrated the importance of PV^+ neurons in gating developmental plasticity[30] and generating gamma oscillations[9,10]. Our results show that these interneurons also have a unique role in visual coding and perception.

METHODS SUMMARY

AAV was injected into the V1 of adult (P40-60) PV-Cre, CAMKIIα-Cre, SOM-Cre or VIP-Cre mice. For recording and behavioural training in awake mice, the head plate was implanted in the same surgery as the virus injection; recording or training was performed 2–6 weeks after surgery. For ChR2 (Arch) activation, an optic fibre coupled to blue (yellow/green) laser was placed on top of the injection site. For measuring orientation tuning, eight repeats of drifting sinusoidal grating (spatial frequency, 0.04 cycles per degree; temporal frequency, 2 Hz; 100% contrast) were presented in 24 directions (0–360°) in pseudorandom sequence. Blocks of trials with or without laser were interleaved. To quantify tuning, we fitted each tuning curve by double-Gaussian function. To test orientation discrimination, mice were trained to discriminate between go and no-go stimuli (drifting gratings at different orientations) for a water reward[20].

Full Methods and any associated references are available in the online version of the paper.

Received 18 May 2011; accepted 11 June 2012.
Published online 8 August 2012.

1. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5,** 793–807 (2004).
2. Xu, X. M., Roby, K. D. & Callaway, E. M. Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *J. Comp. Neurol.* **518,** 389–404 (2010).
3. Ascoli, G. A. *et al.* Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Rev. Neurosci.* **9,** 557–568 (2008).
4. Blanche, T. J., Spacek, M. A., Hetke, J. F. & Swindale, N. V. Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording. *J. Neurophysiol.* **93,** 2987–3000 (2005).
5. Du, J. G. *et al.* High-resolution three-dimensional extracellular recording of neuronal activity with microfabricated electrode arrays. *J. Neurophysiol.* **101,** 1671–1678 (2009).
6. Deisseroth, K. Optogenetics. *Nature Methods* **8,** 26–29 (2011).
7. Chow, B. Y. *et al.* High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463,** 98–102 (2010).
8. Zhang, F. *et al.* Optogenetic interrogation of neural circuits: technology for probing mammalian brain structures. *Nature Protocols* **5,** 439–456 (2010).
9. Sohal, V. S., Zhang, F., Yizhar, O. & Deisseroth, K. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* **459,** 698–702 (2009).
10. Cardin, J. A. *et al.* Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459,** 663–667 (2009).
11. Toledo-Rodriguez, M. *et al.* Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. *Cereb. Cortex* **14,** 1310–1327 (2004).
12. Kerlin, A. M., Andermann, M. L., Berezovskii, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67,** 858–871 (2010).
13. Ma, W. P. *et al.* Visual representations by cortical somatostatin inhibitory neurons–selective but with weak and delayed responses. *J. Neurosci.* **30,** 14371–14379 (2010).
14. Tsien, J. Z. *et al.* Subregion- and cell type-restricted gene knockout in mouse brain. *Cell* **87,** 1317–1326 (1996).
15. Dávid, C., Schleicher, A., Zuschratter, W. & Staiger, J. F. The innervation of parvalbumin-containing interneurons by VIP-immunopositive interneurons in the primary somatosensory cortex of the adult rat. *Eur. J. Neurosci.* **25,** 2329–2340 (2007).
16. Ayaz, A. & Chance, F. S. Gain modulation of neuronal responses by subtractive and divisive mechanisms of inhibition. *J. Neurophysiol.* **101,** 958–968 (2009).
17. Mehaffey, W. H., Doiron, B., Maler, L. & Turner, R. W. Deterministic multiplicative gain control with active dendrites. *J. Neurosci.* **25,** 9968–9977 (2005).
18. Kang, K. J., Shapley, R. M. & Sompolinsky, H. Information tuning of populations of neurons in primary visual cortex. *J. Neurosci.* **24,** 3726–3735 (2004).
19. Schoups, A., Vogels, R., Qian, N. & Orban, G. Practising orientation identification improves orientation coding in V1 neurons. *Nature* **412,** 549–553 (2001).
20. Andermann, M. L. K. e. r. l. i. n. A. M. &. R. e. i. d. R. C. Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front. Cell. Neurosci.* **4,** 3 (2010).
21. Sillito, A. M. Effectiveness of bicuculline as an antagonist of GABA and visually evoked inhibition in the cat's striate cortex. *J. Physiol. (Lond.)* **250,** 287–304 (1975).
22. Nelson, S., Toth, L., Sheth, B. & Sur, M. Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science* **265,** 774–777 (1994).
23. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73,** 159–170 (2012).
24. Chen, Y. *et al.* Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature Neurosci.* **11,** 974–982 (2008).
25. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* **55,** 131–141 (2007).
26. Ferster, D. & Miller, K. D. Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.* **23,** 441–471 (2000).

27. Runyan, C. A. *et al.* Response features of parvalbumin-expressing interneurons suggest precise roles for subtypes of inhibition in visual cortex. *Neuron* **67,** 847–857 (2010).
28. Troyer, T. W., Krukowski, A. E., Priebe, N. J. & Miller, K. D. Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. *J. Neurosci.* **18,** 5908–5927 (1998).
29. Di Cristo, G. *et al.* Subcellular domain-restricted GABAergic innervation in primary visual cortex in the absence of sensory and thalamic inputs. *Nature Neurosci.* **7,** 1184–1186 (2004).
30. Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Rev. Neurosci.* **6,** 877–888 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Goard, L. Pinto, M. Xu and M. Viesel for help with experimental setup, A. Kaluszka for help with data analysis and H. Alitto and M. Poo for discussion and comments on the manuscript. This work was supported by National Institutes of Health grants R01 EY018861 and PN2 EY018241 and National Science Foundation grant 22250400-42533.

Author Contributions S.-H.L. and Y.D. conceived and designed the experiments. S.-H.L. performed and organized all the experiments. A.C.K. developed the head-fixed awake mouse preparation and behavioural task setup. S.Z. performed whole-cell recording experiments. V.P. performed some of the virus injection and head-plate implant surgeries. J.G.F. supported viral vector preparation. S.C.M. manufactured silicon probe (type II). H.T. and Z.J.H. generated SOM-Cre and VIP-Cre mouse lines. E.S.B. generated the Arch–eGFP viral vector. F.Z. and K.D. developed the ChR2–eYFP viral vector. S.-H.L., A.K. and Y.D. analysed the data. S.-H.L. and Y.D. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.D. (ydan@berkeley.edu).

LETTER

doi:10.1038/nature11259

Bacterial virulence proteins as tools to rewire kinase pathways in yeast and immune cells

Ping Wei[1,2]*, Wilson W. Wong[1,2,3]*†, Jason S. Park[1,2,3], Ethan E. Corcoran[2,3,4]†, Sergio G. Peisajovich[1,2]†, James J. Onuffer[1,3], Arthur Weiss[2,3,4] & Wendell A. Lim[1,2,3]

Bacterial pathogens have evolved specific effector proteins that, by interfacing with host kinase signalling pathways, provide a mechanism to evade immune responses during infection[1,2]. Although these effectors contribute to pathogen virulence, we realized that they might also serve as valuable synthetic biology reagents for engineering cellular behaviour. Here we exploit two effector proteins, the *Shigella flexneri* OspF protein[3] and *Yersinia pestis* YopH protein[4], to rewire kinase-mediated responses systematically both in yeast and mammalian immune cells. Bacterial effector proteins can be directed to inhibit specific mitogen-activated protein kinase pathways selectively in yeast by artificially targeting them to pathway-specific complexes. Moreover, we show that unique properties of the effectors generate new pathway behaviours: OspF, which irreversibly inactivates mitogen-activated protein kinases[4], was used to construct a synthetic feedback circuit that shows novel frequency-dependent input filtering. Finally, we show that effectors can be used in T cells, either as feedback modulators to tune the T-cell response amplitude precisely, or as an inducible pause switch that can temporarily disable T-cell activation. These studies demonstrate how pathogens could provide a rich toolkit of parts to engineer cells for therapeutic or biotechnological applications.

Many bacterial pathogens have developed an array of effector proteins to rewire host signalling networks and downregulate the immune response[2] (Fig. 1a). Some effectors mimic host activities, such as the *Y. pestis* effector YopH, which is a highly active phosphotyrosine phosphatase[5]. Other effectors use unusual mechanisms, such as the *S. flexneri* OspF protein, which irreversibly inactivates mitogen-activated protein kinases (MAPKs) by catalysing a β-elimination reaction that removes the hydroxyl group of the key phosphothreonine side chain[4].

MAPK pathways have a central role in diverse eukaryotic responses, ranging from immune response to cell-fate decisions[5,6]. Thus, the ability to tune MAPK response would facilitate engineering cells for diverse therapeutic and biotechnological applications[7,8]. Recent work has shown that MAPK signalling dynamics in yeast can be reshaped with synthetic feedback loops that involve controlled expression and targeting of pathway modulators to appropriate signalling complexes[9]. Identifying effective pathway modulators is challenging; thus we considered that pathogen effector proteins may have untapped use as components for predictably and systematically engineering signalling pathways. Here we use the effector proteins OspF and YopH to modulate kinase signalling pathways in yeast and in human primary T cells.

We first introduced OspF into yeast. As reported[10], overexpression of OspF led to growth inhibition under standard conditions, hyperosmotic stress conditions (Fig. 1b) and cell-wall-damaging conditions (Supplementary Fig. 1a). OspF contains a canonical docking peptide at its amino (N) terminus that allows it to bind several MAPKs in yeast[11]. We found that expression of an OspF mutant lacking its native docking peptide (ΔN-OspF) yielded normal growth behaviour under all conditions (Fig. 1b and Supplementary Fig. 1a). Next we tested whether ΔN-OspF could be redirected to a specific pathway by tagging the protein with a leucine zipper heterodimerization motif, and fusing the complementary interacting motif to Pbs2, the scaffold protein that organizes the osmolarity MAPK pathway. This targeted version of ΔN-OspF only showed a growth defect under high salt conditions, showing that OspF activity could be engineered to inhibit a specific MAPK (Fig. 1b).

To explore re-targeting OspF to specific pathways further, we engineered yeast strains in which ΔN-OspF was selectively targeted to either the osmolarity MAPK complex or the mating MAPK complex (by targeting it to the mating pathway scaffold protein, Ste5) (Fig. 1c). Targeting of ΔN-OspF to the Pbs2 inhibited the osmolarity response but not the mating response. Conversely, when ΔN-OspF was targeted to Ste5, only the mating response was inhibited. Thus, the inhibitory activity of this effector could be selectively aimed at one of several MAPK pathways in the same cell.

One of the unique aspects of OspF is that it catalyses an irreversible inactivation of MAPKs (unlike reversible dephosphorylation by a phosphatase). Thus MAPK activity can only be restored through new protein synthesis, which has a much slower timescale than re-phosphorylation (Fig. 2a, b). This longer timescale would be expected to lead to an extended refractory period after OspF action, during which the targeted MAPK pathway could no longer respond to subsequent stimuli.

Computational simulations indicated that a long refractory period could result in significant changes to the frequency-dependent behaviour of pathway response (Supplementary Fig. 2). There is growing evidence that cells use frequency modulation of diverse molecular events to encode and transmit information[12,13]. Our models indicated that with a negative feedback loop (that is, MAPK activity induced expression of OspF), pathway output would be dampened when input periods are long enough to accumulate significant amounts of the negative effector but shorter than the refractory period (Supplementary Fig. 2).

To test if OspF could be used to filter frequency dependent inputs, we constructed a synthetic negative feedback loop in the yeast osmo-response pathway by expressing OspF targeted to the osmo-response signalling complex (ΔN-OspF–zipper) from the Hog1 responsive promoter, *pSTL1* (Fig. 2c). As a comparison, we also engineered an analogous synthetic feedback loop using a reversible Hog1 MAPK inhibitory protein—the yeast MAPK phosphatase, PTP2. Phospho-Hog1 translocation to the nucleus was used as a fast-timescale output reporter (Supplementary Fig. 3)[14–16]. To measure integrated output over a longer timescale, we also measured a slower timescale

[1]Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, California 94158, USA. [2]Howard Hughes Medical Institute, University of California San Francisco, San Francisco, California 94158, USA. [3]The Cell Propulsion Laboratory, University of California San Francisco, San Francisco, California 94158, USA. [4]Department of Medicine, University of California San Francisco, San Francisco, California 94158, USA. †Present addresses: Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA (W.W.W.); Department of Pulmonology and Critical Care Medicine, Northwest Permanente P.C., 10180 SE Sunnyside Road, Clackamas, Oregon 97015, USA (E.E.C.); Department of Cell & Systems Biology, University of Toronto, Toronto, Ontario M5S 3G5, Canada (S.G.P.).

*These authors contributed equally to this work.

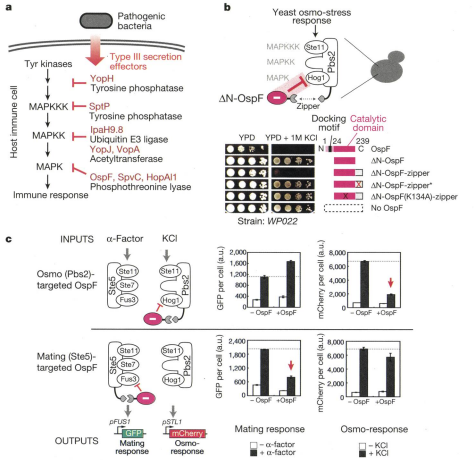


Figure 1 | Bacterial effector OspF can block selective MAPK pathways in yeast. a, Type III secretion effectors that modulate host kinase signalling. MAPKKK, MAP-kinase kinase kinase. **b**, Targeting of OspF to yeast osmolarity pathway. Wild-type OspF impairs growth on rich media, but is rescued by docking motif deletion (ΔN-OspF). Recruitment of ΔN-OspF to osmolarity scaffold Pbs2 by leucine zipper selectively blocks growth on 1 M KCl (zipper*, mutant leucine zipper; K134A, catalytic dead mutant of OspF). **c**, ΔN-OspF selectively inhibits mating or osmolarity if targeted to appropriate scaffold complex, assayed using pathway-specific transcriptional reporters. Average fluorescence and s.d. of three experiments are shown. a.u., arbitrary units.

transcriptional reporter—expression of mCherry from the *pSTL1* promoter.

We found that the OspF-mediated negative feedback circuit altered the osmo-stress pathway response to intermediate frequency stimulation, but not to continuous stress or to high-frequency stimulation (Fig. 2d and Supplementary Fig. 4a). Furthermore, in the course of stimulating cells with pulses of KCl of varying length, we discovered that an input period of approximately 16 min (intermediate frequency) leads to highly divergent transcriptional responses (Fig. 2d). Examination of Hog1 nuclear import in the OspF feedback strain shows that after three pulses, the amount of Hog1 competent for nuclear localization has decreased to near zero, consistent with a model in which—in this timeframe—there is sufficient OspF to inactivate the bulk of the Hog1 population and now to render the cells refractory to further pulses of stimulation. The cells containing a PTP2 negative feedback circuit do not show this marked filtering at this frequency. With higher frequency stimulation (2 min period), the three strains also do not significantly differ from each other in response, presumably because there is inadequate activation time with each pulse to build up a sufficient concentration of effector[17].

A broad frequency dependence analysis shows that the wild-type osmo-response pathway functions as a band-pass filter, with maximal response at intermediate frequencies, whereas the engineered pathway more closely resembles a low-pass filter, with maximal responses at low frequency (Fig. 2e). These distinct frequency filtering behaviours fit those predicted by computational simulations (Supplementary Fig. 2d).

We then sought to test whether these bacterial effectors could be used to rewire signalling in immune cells. Human T cells are an attractive synthetic biology platform because they can be isolated from patients, genetically engineered *ex vivo* and then transferred back into patients to treat cancer and chronic infection[18,19]. Although promising, the therapeutic application of engineered T cells carries risks of adverse side effects including inadvertent autoimmune-like attack of off-target host tissues[20,21]. Thus mechanisms to control the specificity, amplitude and timing of T-cell function are critical to balance therapeutic action against off-target toxicity.

Both OspF and YopH can modify the T-cell receptor (TCR) pathway (Fig. 3a). OspF inactivates the MAPK extracellular signal-regulated kinase (ERK), which is a central component of TCR signalling[4,22], whereas YopH dephosphorylates phospho-tyrosine, including the T-cell scaffold proteins LAT and SLP-76 (ref. 23). Constitutive expression of YopH and OspF in *Jurkat* T cells leads to severe inhibition of TCR activation, as measured by an NFAT transcriptional reporter[24] (Fig. 3a) (as well as other reporters of T-cell activation; Supplementary Fig. 5a). Expression of the catalytic dead versions of YopH and OspF had no effect on the TCR activation (Supplementary Fig. 5b). In addition, we showed that these two effectors clearly target distinct steps of the T-cell activation pathway, because induction of the T cells with the combination of phorbol 12-myristate 13-acetate (PMA) and ionomycin, which activates the T cell downstream of the PLCγ1-LAT/SLP76 dependent response, bypassed YopH inhibition[25], but was sensitive to OspF inhibition (Fig. 3a and Supplementary Fig. 5a). Thus, distinct pathogen effector proteins can be used to block this

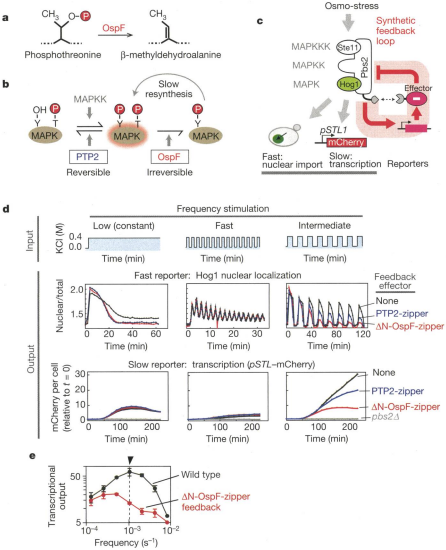


Figure 2 | Tuning-frequency-dependent response of yeast osmolarity pathway using synthetic OspF feedback loop. a, OspF is MAPK phosphothreonine lyase. **b**, OspF irreversibly inhibits MAPK activity. **c**, Synthetic negative feedback loop was built by expressing OspF from osmo-inducible *pSTL1* promoter. Hog1–GFP nuclear accumulation (Nuc:Tot ratio) was measured as a fast output reporter. *pSTL1*–mCherry served as a slower, gene expression reporter. **d**, Cells were stimulated with variable frequency inputs: low (constant), high (period = 2 min) and intermediate (period = 16 min). Significant difference in response to intermediate frequency stimulation is observed. **e**, Frequency–response curves for wild-type and OspF feedback strains (arrow: period = 16 min). Each point is average of 50–100 cells; s.d. from three repeats is shown. More detail on the frequency analysis is given in Supplementary Figs 2–4.

pathway at particular steps, much like a specific small-molecule inhibitor.

Given the ability of OspF and YopH to modulate T-cell responses, we sought to use these proteins to build circuits that could, in principle, improve the safety of therapeutic T cells. In adoptive T-cell therapy, a challenge is to limit over-activation or off-target activation of T cells that could lead to killing of host cells or to cytokine storm—a life-threatening immune response. One approach is to incorporate a safety 'kill switch'[26–28] into the T cells, such as the herpes simplex virus thymidine kinase gene. This protein converts the pro-drug ganciclovir into an inhibitor of replication, thus killing cells expressing the gene. Although herpes simplex virus thymidine kinase is currently being tested in a phase III clinical trial for the treatment of graft versus host disease in bone marrow transplants, this strategy irreversibly destroys the engineered, adoptively transferred cells[29]. Thus, instead of killing the engineered cells, we sought to design circuits that would limit the amplitude of the T-cell response or to pause T-cell activity temporarily.

We first tested whether bacterial effectors could be used to limit the response amplitude of *Jurkat* T cells. Negative feedback loops can act to limit the maximal amplitude of a response[30], so we engineered a library of negative feedback loops in which the OspF and YopH were expressed from a series of TCR-responsive promoters of varying strength (AP1 and NFAT) (Fig. 3b and Supplementary Fig. 7). For further tuning of feedback parameters, we also tagged effectors with degradation sequences (PEST motif) that reduce the half-life of the effectors. This series of negative feedback loops led to controlled reduction of the maximal response amplitude of T-cell activation (Fig. 3b). Moreover, the amplitude could be tuned systematically by varying feedback promoter strength and effector stability (Supplementary Fig. 7b). For example, expression of OspF from the strong feedback (*pNFAT*) promoter leads to a very low maximal response amplitude, but this effect could be systematically tuned by destabilizing the OspF effector with a PEST tag.

We also tested whether the bacterial effectors could be used to construct pause switches, which could transiently and reversibly disable T cells. We placed the effectors under the control of a tetracycline inducible promoter (*pTRE*), which allowed external control of the timing of effector expression with the addition of doxycycline.

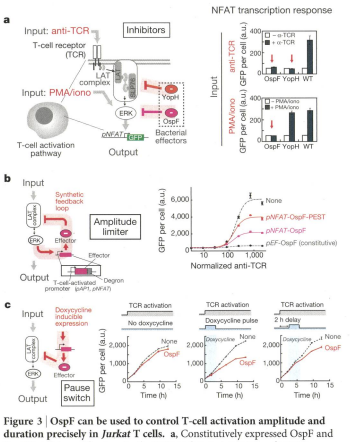


Figure 3 | OspF can be used to control T-cell activation amplitude and duration precisely in *Jurkat* T cells. **a**, Constitutively expressed OspF and YopH inhibit specific steps in the TCR pathway. Activation of NFAT transcription by anti-TCR antibody versus PMA/ionomycin is shown. **b**, Synthetic amplitude limiters were constructed using negative feedback loops with different effectors, promoters and degrons. Dose–response curves of the *pNFAT*-OspF circuit, with or without degron, are shown. **c**, Synthetic pause switch was built by expressing OspF from a doxycycline-inducible promoter, *pTRE*. A 4 h pulse induction disabled T-cell response when initiated at 0 or 2 h after T-cell stimulation. Standard deviation from three samples is shown.

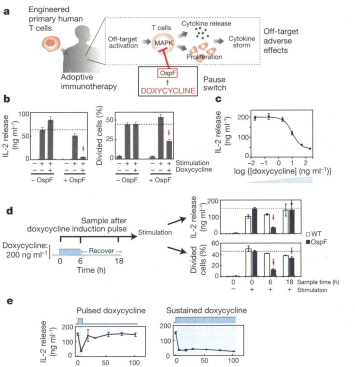


Figure 4 | OspF can be used as a synthetic pause switch to control human primary CD4$^+$ T-cell activation. **a**, Off-target activation of transplanted T cells can induce cytokine storm in adoptive immunotherapy. A pause switch could prevent this adverse response. **b**, Six-hour pre-induction of OspF by doxycycline inhibited IL-2 release and proliferation of activated CD4$^+$ T cells. **c**, Six-hour pre-treatment with different doses of doxycycline can tune IL-2 release. **d**, Cells treated with 6 h doxycycline pulse were sampled at different times after the pulse, then subjected to a 24 h IL-2 release assay or a 4 day proliferation assay. **e**, IL-2 release can be inhibited either transiently or in a sustained manner by varying the duration of doxycycline treatment. Average and s.d. of three experimental repeats are shown.

Effectors were fused to a destabilization domain so that they would be rapidly degraded once doxycycline is removed. Using this system, we first showed that transient expression of bacterial effectors can inhibit TCR signalling after the pathway is activated in *Jurkat* T cells (Fig. 3c). TCR signalling can be inhibited up to 6 h after activation using this approach (Supplementary Fig. 8b). Finally, we showed that engineered T cells can be subjected to cycles of TCR activation, paused with a short period of induced expression of the bacterial effector, and then reactivated after this pause (Supplementary Fig. 8c).

We then tested the pause switch in a clinically important cell type for adoptive immunotherapy—primary human CD4$^+$ T cells (in contrast to the *Jurkat* T-cell line, which does not require cytokine or TCR activation to stimulate proliferation). We showed that when OspF is induced by the addition of doxycycline, both interleukin-2 (IL-2) release and cell proliferation were inhibited in a dose-dependent manner (Fig. 4b, c). Activation of the TCR by anti-CD3/CD28 and antigen-presenting cells can also be inhibited by expression of OspF (Supplementary Fig. 9a). Moreover, after doxycycline is removed, IL-2 release and cell division recovers in 6–18 h (Fig. 4d, e). Sustained exposure to doxycycline can inhibit T-cell activity over the course of several days (Fig. 4e) without having any significant effect on cell viability (Supplementary Fig. 9c). Thus this work provides a proof of principle for the design of a simple 'pause' switch that could allow external control over the timing and level of T-cell activation and cytokine release, to minimize adverse events associated with adoptive immunotherapy such as cytokine storm.

Most work on bacterial pathogen effector proteins has the long-term aim of neutralizing the pathogens' infectious capabilities. We have shown, however, that bacterial effectors can also be valuable synthetic biology tools, because of their unique biochemical properties. We have used bacterial effectors to modulate MAPK signalling in yeast to generate novel time-dependent dynamics. We have also shown that bacterial effectors can be used to tune human TCR signalling dynamics flexibly, with potential application as safety switches for adoptive immunotherapy. The vast array of bacterial pathogen effector proteins, beyond those studied here, holds promise as a rich and important source of parts for the cellular engineering toolkit.

METHODS SUMMARY

Flow cytometry experiments. Analysis of the pathway-dependent fluorescent protein (green fluorescent protein (GFP) and mCherry) expression in yeast cells and phosphorylated ERK in *Jurkat* T cells was performed with a BD LSRII flow cytometer (BD Biosciences) equipped with a high-throughput sampler. α-Factor (1.5 μM; GenScript) or 0.4 M of KCl were added into yeast cultures to induce separately the mating- or osmo-specific pathway response. For staining of phosphorylated ERK in *Jurkat* T cells, cells were fixed, made permeable by incubation with ice-cold 90% methanol on ice for 30 min and stained with primary antibody to phosphorylated ERK (Cell Signaling) and anti-rabbit APC secondary antibody (Jackson Immunoresearch).

Microfluidics and fluorescent microscopy. Microfluidic yeast cell culture was performed in Y04C plates with an ONIX flow control system (Cellasic). Cells were loaded into the flow chamber pre-coated with concanavalin A. Pulse stimulation with salt media was performed by ONIX FG flow control software (Cellasic) with a flow pressure of 8 pounds per square inch. Image acquisition was performed with a TE2000-E automated inverted microscope (Nikon) with Perfect Focus and a ×100

oil immersion lens. Image analysis for both nuclear Hog1–GFP import and *pSTL1*–mCherry expression was performed with custom MATLAB (Mathworks) software.

Human T-cell activation assay. Resting human primary $CD4^+$ cells (transduced with the pause switch constructs or untransduced) were pretreated with 200 ng ml^{-1} doxycycline for 6 h. Fifty thousand cells were placed in a 96-well plate with 200 μl human growth media with activation agents added (10 ng ml^{-1} PMA + 0.5 μM ionomycin, magnetic Dynabeads coated with anti-CD3/anti-CD28 (beads:cells ratio, 0.3:1)). After 24 h of incubation at 37 °C, the released IL-2 in the supernatant was measured with a human IL-2 ELISA kit II (BD Biosciences). Cells labelled with CellTrace Violet dye (Invitrogen) were assayed by flow cytometry after incubation for 4 days to quantify cell proliferation.

Full Methods and any associated references are available in the online version of the paper.

Received 21 March; accepted 25 May 2012.
Published online 22 July 2012.

1. Ribet, D. & Cossart, P. Pathogen-mediated posttranslational modifications: a re-emerging field. *Cell* **143,** 694–702 (2010).
2. Broberg, C. A. & Orth, K. Tipping the balance by manipulating post-translational modifications. *Curr. Opin. Microbiol.* **13,** 34–40 (2010).
3. Zhang, Z. Y. *et al.* Expression, purification, and physicochemical characterization of a recombinant *Yersinia* protein tyrosine phosphatase. *J. Biol. Chem.* **267,** 23759–23766 (1992).
4. Li, H. *et al.* The phosphothreonine lyase activity of a bacterial type III effector family. *Science* **315,** 1000–1003 (2007).
5. Shaw, A. S. & Filbert, E. L. Scaffold proteins and immune-cell signalling. *Nature Rev. Immunol.* **9,** 47–56 (2009).
6. Dong, C., Davis, R. J. & Flavell, R. A. MAP kinases in the immune response. *Annu. Rev. Immunol.* **20,** 55–72 (2002).
7. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nature Rev. Genet.* **11,** 367–379 (2010).
8. Purnick, P. E. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature Rev. Mol. Cell Biol.* **10,** 410–422 (2009).
9. Bashor, C. J., Helman, N. C., Yan, S. & Lim, W. A. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* **319,** 1539–1543 (2008).
10. Kramer, R. W. *et al.* Yeast functional genomic screens lead to identification of a role for a bacterial effector in innate immunity regulation. *PLoS Pathog.* **3,** e21 (2007).
11. Zhu, Y. *et al.* Structural insights into the enzymatic mechanism of the pathogenic MAPK phosphothreonine lyase. *Mol. Cell* **28,** 899–913 (2007).
12. Nelson, D. E. *et al.* Oscillations in NF-κB signaling control the dynamics of gene expression. *Science* **306,** 704–708 (2004).
13. Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455,** 485–490 (2008).
14. Muzzey, D., Gomez-Uribe, C. A., Mettetal, J. T. & van Oudenaarden, A. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* **138,** 160–171 (2009).
15. Mettetal, J. T., Muzzey, D., Gomez-Uribe, C. & van Oudenaarden, A. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Science* **319,** 482–484 (2008).
16. Hersen, P., McClean, M. N., Mahadevan, L. & Ramanathan, S. Signal processing by the HOG MAP kinase pathway. *Proc. Natl Acad. Sci. USA* **105,** 7165–7170 (2008).
17. Pelet, S. *et al.* Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* **332,** 732–735 (2011).
18. Morgan, R. A., Dudley, M. E. & Rosenberg, S. A. Adoptive cell therapy: genetic modification to redirect effector cell specificity. *Cancer J.* **16,** 336–341 (2010).
19. June, C. H., Blazar, B. R. & Riley, J. L. Engineering lymphocyte subsets: tools, trials and tribulations. *Nature Rev. Immunol.* **9,** 704–716 (2009).
20. Morgan, R. A. *et al.* Case report of a serious adverse event following the administration of T cells transduced with a chimeric antigen receptor recognizing ERBB2. *Mol. Ther.* **18,** 843–851 (2010).
21. Brentjens, R., Yeh, R., Bernal, Y., Riviere, I. & Sadelain, M. Treatment of chronic lymphocytic leukemia with genetically targeted autologous T cells: case report of an unforeseen adverse event in a phase I clinical trial. *Mol. Ther.* **18,** 666–668 (2010).
22. Arbibe, L. *et al.* An injected bacterial effector targets chromatin access for transcription factor NF-κB to alter transcription of host genes involved in immune responses. *Nature Immunol.* **8,** 47–56 (2007).
23. Gerke, C., Falkow, S. & Chien, Y. H. The adaptor molecules LAT and SLP-76 are specifically targeted by *Yersinia* to inhibit T cell activation. *J. Exp. Med.* **201,** 361–371 (2005).
24. Lin, J. & Weiss, A. The tyrosine phosphatase CD148 is excluded from the immunologic synapse and down-regulates prolonged T cell signaling. *J. Cell Biol.* **162,** 673–682 (2003).
25. Yao, T., Mecsas, J., Healy, J. I., Falkow, S. & Chien, Y. Suppression of T and B lymphocyte activation by a *Yersinia pseudotuberculosis* virulence factor, yopH. *J. Exp. Med.* **190,** 1343–1350 (1999).
26. Kieback, E., Charo, J., Sommermeyer, D., Blankenstein, T. & Uckert, W. A safeguard eliminates T cell receptor gene-modified autoreactive T cells after adoptive transfer. *Proc. Natl Acad. Sci. USA* **105,** 623–628 (2008).
27. de Witte, M. A. *et al.* An inducible caspase 9 safety switch can halt cell therapy-induced autoimmune disease. *J. Immunol.* **180,** 6365–6373 (2008).
28. Bonini, C. *et al.* The suicide gene therapy challenge: how to improve a successful gene therapy approach. *Mol. Ther.* **15,** 1248–1252 (2007).
29. Ciceri, F. *et al.* Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haemopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I–II study. *Lancet Oncol.* **10,** 489–500 (2009).
30. Brandman, O. *et al.* Feedback loops shape cellular signals in space and time. *Science* **322,** 390–395 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Orth for the YopH plasmid; K. McNally and S. Neou for tissue culture support; H. El-Samad, C. Voigt, C. Tang and the Lim laboratory for discussions. We acknowledge the 2007 UCSF iGEM team (M. Chen, E. Chou, J. Huang, L. Jann, E. Meltzer, A. Ng and R. Ovadia) for their initial work on bacterial effectors in yeast. This work was supported by American Cancer Society fellowship PF-09-137-01-TBE (W.W.), a Li Foundation Fellowship (P.W.), a California Institute for Regenerative Medicine fellowship (grant number TG2-01153) (J.S.P.), National Institutes of Health grants PN2EY016546, RO1GM055040, RO1GM062583 and P50GM081879 (W.A.L.), the NSF Synthetic Biology and Engineering Research Center (W.A.L.), the Packard Foundation (W.A.L.) and the Howard Hughes Medical Institute (A.W. and W.A.L.).

Author Contributions P.W., S.G.P. and W.A.L. initiated the project in yeast. W.W., E.E.C., A.W. and W.A.L. initiated the project in T cells. P.W., W.W.W. and W.A.L. wrote the paper. P.W. and S.G.P. planned and performed the experiments in yeast. W.W.W., P.W., E.E.C., J.J.O. and J.S.P. planned and performed the experiment in T cells.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.A.L. (lim@cmp.ucsf.edu).

LETTER

doi:10.1038/nature11250

NLRP6 negatively regulates innate immunity and host defence against bacterial pathogens

Paras K. Anand[1], R. K. Subbarao Malireddi[1], John R. Lukens[1], Peter Vogel[2], John Bertin[3], Mohamed Lamkanfi[4,5] & Thirumala-Devi Kanneganti[1]

Members of the intracellular nucleotide-binding and oligomerization domain (NOD)-like receptor (NLR) family contribute to immune responses through activation of nuclear factor-κB (NF-κB), type I interferon and inflammasome signalling[1]. Mice lacking the NLR family member NLRP6 were recently shown to be susceptible to colitis and colorectal tumorigenesis[2–4], but the role of NLRP6 in microbial infections and the nature of the inflammatory signalling pathways regulated by NLRP6 remain unclear. Here we show that *Nlrp6*-deficient mice are highly resistant to infection with the bacterial pathogens *Listeria monocytogenes, Salmonella typhimurium* and *Escherichia coli*. Infected *Nlrp6*-deficient mice had increased numbers of monocytes and neutrophils in circulation, and NLRP6 signalling in both haematopoietic and radioresistant cells contributed to increased susceptibility. *Nlrp6* deficiency enhanced activation of mitogen-activated protein kinase (MAPK) and the canonical NF-κB pathway after Toll-like receptor ligation, but not cytosolic NOD1/2 ligation, *in vitro*. Consequently, infected *Nlrp6*-deficient cells produced increased levels of NF-κB- and MAPK-dependent cytokines and chemokines. Thus, our results reveal NLRP6 as a negative regulator of inflammatory signalling, and demonstrate a role for this NLR in impeding clearance of both Gram-positive and -negative bacterial pathogens.

Despite the availability of antibiotics, bacterial infections continue to threaten public health worldwide. According to estimates from the US Centers for Disease Control and Prevention, roughly 48 million people in the United States get ill every year owing to foodborne illnesses, resulting in 3,000 deaths annually[5]. *Listeria* and *Salmonella* belong to the top three pathogens contributing to foodborne infections resulting in death[5,6]. *L. monocytogenes* and *S. typhimurium* are facultative intracellular pathogens that can survive and replicate within macrophages and dendritic cells. *Listeria* rapidly escapes the phagosome to replicate in the cytosol, whereas *Salmonella* lacks a phagosomal escape mechanism and inhabits the *Salmonella*-containing vacuole[7–9]. A deeper understanding of the mechanisms by which these and other bacterial pathogens are sensed by the immune system may contribute to new approaches for developing antimicrobials. Cells of the innate immune system detect microorganisms by means of a limited set of evolutionary conserved pattern recognition receptors (PRRs)[10]. NLR proteins represent a family of intracellular PRRs that survey the cytoplasmic compartment for infectious agents and cellular damage[1]. NLRs such as NOD1 and NOD2 contribute to host defence against microbial pathogens by inducing the production of pro-inflammatory cytokines through activation of NF-κB and MAPK signalling, whereas NLRs such as NLRP1, NLRP3 and NLRC4 facilitate the activation of inflammatory caspases in large multiprotein complexes termed inflammasomes[11,12]. Our knowledge of the former NLRs has markedly improved in recent years, but initial characterization of NLRP6 has only recently been reported[2–4,13]. These studies demonstrated that NLRP6 contributes to protection against colitis, colorectal tumorigenesis and non-alcoholic steatosis by regulating the integrity of the epithelial barrier and by altering the composition of the gut microflora[2–4,13]. However, the role of NLRP6 in host defence against microbial pathogens, and the signalling pathways it regulates remain unclear.

To characterize NLRP6 expression in immune cells, *Nlrp6* transcript abundance in immune and epithelial cells was analysed by real-time quantitative PCR (qPCR). As previously reported[14], neutrophils and T cells showed the highest expression levels for *Nlrp6*, followed by macrophages, epithelial cells and dendritic cells (Supplementary Fig. 1), suggesting that NLRP6 may have an important role in these cell types. The domain architecture of NLRP6 resembles that of NLRP3, and consists of an amino-terminal Pyrin domain, a centrally located nucleotide-binding domain and a stretch of carboxy-terminally located leucine-rich repeats[1]. *Nlrp6*-deficient mice have been described[2–4,13], and animals appeared healthy and did not display gross abnormalities when housed in a specific-pathogen-free facility (data not shown). To characterize the role of NLRP6 during microbial infections, wild-type (C57BL/6J) and $Nlrp6^{-/-}$ mice were infected intraperitoneally (i.p.) with a lethal dose of 10^6 colony-forming units (c.f.u.) of *L. monocytogenes*. Whereas the entire cohort of wild-type mice succumbed to infection within 5–6 days, 75% of $Nlrp6^{-/-}$ mice survived the infection and were still alive at day 20 after infection (Fig. 1a). In agreement, wild-type mice lost on average 20% of their initial body weight before succumbing to infection, whereas in $Nlrp6^{-/-}$ mice weight loss was initially limited (<10% by day 3) and body weight gradually returned to normal levels after day 4 (Fig. 1b). To examine whether differential mortality in wild-type and $Nlrp6^{-/-}$ mice was associated with differences in bacterial dissemination, bacterial burdens in systemic organs were determined at days 1 and 3 after infection. Notably, both the liver and spleen of $Nlrp6^{-/-}$ mice contained significantly less bacteria than those of wild-type mice at day 1 (Fig. 1c, d); this difference increased further by day 3 after infection (Fig. 1e, f). Immunohistochemical analysis of liver sections confirmed *L. monocytogenes* to be nearly absent from the liver capsules of $Nlrp6^{-/-}$ mice, whereas those of wild-type mice were morphologically distorted owing to high pathogen counts (Fig. 1g). *L. monocytogenes* infection is known to trigger histopathological lesions and the formation of inflammatory cell foci, the size of which correlates with disease severity[15]. To examine the extent of immune cell infiltration, haematoxylin and eosin (H&E)-stained liver sections of *L. monocytogenes*-infected wild-type and $Nlrp6^{-/-}$ mice were compared. In agreement with our previous results, inflammatory cell foci in $Nlrp6^{-/-}$ livers were significantly smaller than those found in wild-type mice at day 3 after infection (Fig. 1h and Supplementary Fig. 2). To examine the role of NLRP6 during alternative infection routes, cohorts of wild-type and $Nlrp6^{-/-}$ mice were infected with *L. monocytogenes* intravenously (i.v.). As with i.p.-infected animals (Fig. 1e, f), bacterial counts in the liver and spleen of i.v.-infected $Nlrp6^{-/-}$ mice were significantly lower than

[1]Department of Immunology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [2]Animal Resources Center and the Veterinary Pathology Core, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [3]Pattern Recognition Receptor DPU, Immuno-Inflammation TA, GlaxoSmithKline, Collegeville, Pennsylvania 19426, USA. [4]Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium. [5]Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium.

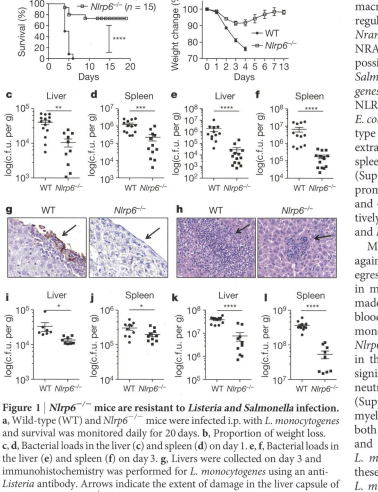


Figure 1 | *Nlrp6*$^{-/-}$ mice are resistant to *Listeria and Salmonella* infection. **a**, Wild-type (WT) and *Nlrp6*$^{-/-}$ mice were infected i.p. with *L. monocytogenes* and survival was monitored daily for 20 days. **b**, Proportion of weight loss. **c**, **d**, Bacterial loads in the liver (**c**) and spleen (**d**) on day 1. **e**, **f**, Bacterial loads in the liver (**e**) and spleen (**f**) on day 3. **g**, Livers were collected on day 3 and immunohistochemistry was performed for *L. monocytogenes* using an anti-*Listeria* antibody. Arrows indicate the extent of damage in the liver capsule of wild-type and *Nlrp6*$^{-/-}$ mice. **h**, Livers from *L. monocytogenes*-infected mice showing inflammatory lesions indicated by arrows. Original magnifications for all images, ×40. **i**, **j**, Wild-type and *Nlrp6*$^{-/-}$ mice were infected i.p. with *S. typhimurium*. Bacterial loads were determined in the liver (**i**) and spleen (**j**) on day 1. **k**, **l**, Bacterial loads in the liver (**k**) and spleen (**l**) on day 3. Each point represents an individual mouse and the mean ± s.e.m. are shown. Results show cumulative data from two different experiments. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

the burdens measured in wild-type mice at day 3 after infection (Supplementary Fig. 3). *Nlrp6*$^{-/-}$ mice were recently shown to have an altered microflora composition that is transferable to co-housed wild-type mice[3]. In agreement, 16S ribosomal RNA analysis confirmed an increased abundance of Bacteroidetes (Prevotellaceae family) in the gastrointestinal tract of *Nlrp6*$^{-/-}$ mice relative to the microbiota composition of separately housed wild-type mice (Supplementary Fig. 4a). Co-housing wild-type and *Nlrp6*$^{-/-}$ mice for 4 weeks equalized the prevalence of Prevotellaceae in these mice (Supplementary Fig. 4a, b). However, co-housing did not alter the resistant phenotype of *Nlrp6*$^{-/-}$ mice to *L. monocytogenes* infection because *Nlrp6*$^{-/-}$ mice contained fewer bacteria in the spleen and liver compared with the levels of co-housed wild-type mice (Supplementary Fig. 4c, d). These results suggest that NLRP6 regulates *L. monocytogenes* infection independently of its microflora composition.

To determine whether resistance of *Nlrp6*$^{-/-}$ mice to infection is specific to *L. monocytogenes* or extends to other pathogens, cohorts of wild-type and *Nlrp6*$^{-/-}$ mice were infected i.p. with *S. typhimurium*. Resistance to *S. typhimurium* infection was increased in *Nlrp6*$^{-/-}$ mice as illustrated by the significantly lower bacterial burdens detected in the liver and spleen at days 1 (Fig. 1i, j) and 3 (Fig. 1k, l) after infection. Inbred mice may carry distinct alleles of mouse *Nramp1* (also known as *Slc11a1*, solute carrier family 11 and natural resistance-associated macrophage protein one), the gene products of which may differentially regulate host resistance to *Salmonella* and other bacterial pathogens[16]. *Nramp1* transcripts of both *Nlrp6*$^{-/-}$ and wild-type mice encoded the NRAMP1(Gly169Asp) polymorphism (data not shown), ruling out a possible effect on the increased resistance of *Nlrp6*$^{-/-}$ mice to *Salmonella* infection. Because both *S. typhimurium* and *L. monocytogenes* are facultative intracellular pathogens, we determined whether NLRP6 modulated resistance against extracellular pathogens such as *E. coli* as well. To this end, we analysed c.f.u. in systemic organs of wild-type and *Nlrp6*$^{-/-}$ mice 2 days after infection with this Gram-negative extracellular pathogen. Notably, bacterial burdens in the liver and spleen of *Nlrp6*$^{-/-}$ mice were significantly lower than in wild-type mice (Supplementary Fig. 5). Together, these results suggest that NLRP6 may promote systemic dissemination and growth of both Gram-negative and -positive bacteria (*S. typhimurium* and *L. monocytogenes*, respectively) as well as intracellular and extracellular bacteria (*S. typhimurium* and *E. coli*, respectively) in infected hosts.

Myeloid cells have important roles in early host defence responses against microbial pathogens. They develop in the bone marrow, but egress into the circulation after infection[17]. Because *Nlrp6* is expressed in monocytes and neutrophils (Supplementary Fig. 1), we initially made use of an automated haematology analyser—which determines blood cell types based on morphology—to characterize the influx of monocytes and neutrophils in peripheral blood of wild-type and *Nlrp6*$^{-/-}$ mice. Before infection, monocyte and neutrophil precursors in the bone marrow were similar (Supplementary Fig. 6), and no significant differences in the counts of circulating monocytes and neutrophils were noted in naive wild-type and *Nlrp6*$^{-/-}$ mice (Supplementary Fig. 7), suggesting that basal differentiation of myeloid progenitor cells was not affected in these mice. However, both the fraction and total cell counts of circulating monocytes and neutrophils were markedly higher in the peripheral blood of *L. monocytogenes*-infected *Nlrp6*$^{-/-}$ mice (Fig. 2a, b). To confirm these results, immune cell infiltration in the peritoneal cavity of *L. monocytogenes*-infected animals was analysed by flow cytometry. Basal monocyte and neutrophil populations in the peritoneal cavity of naive wild-type and *Nlrp6*$^{-/-}$ mice did not differ (data not shown). By contrast, significantly higher infiltration of monocytes and neutrophils was observed in the peritoneal cavity of *L. monocytogenes*-infected *Nlrp6*$^{-/-}$ mice compared with wild-type mice (Fig. 2c, d and Supplementary Fig. 8a, b). Unlike granulocyte populations, the numbers of circulating lymphocytes and infiltrated T cells in the peritoneal cavity were similar in *L. monocytogenes*-infected wild-type and *Nlrp6*$^{-/-}$ mice (Supplementary Fig. 8c, d). Concurrently, *Nlrp6* deficiency did not alter the induction of T-cell responses because the frequency and number of ovalbumin (OVA)-specific CD8^{+} T cells were similar in wild-type and *Nlrp6*$^{-/-}$ mice that were infected with OVA-expressing *Listeria* (Supplementary Fig. 9). Taken together, these results suggest that NLRP6 inhibits the influx of monocytes and neutrophils to the circulation and to the peritoneum during bacterial infections.

To examine further the cell types responsible for NLRP6-mediated dissemination of bacterial infections, we created wild-type and *Nlrp6*$^{-/-}$ bone marrow chimaeras. Bone marrow reconstitution was confirmed to reach a level of approximately 95% in irradiated mice (data not shown). As expected, *Nlrp6*$^{-/-}$ mice that were transplanted with *Nlrp6*$^{-/-}$ bone marrow were more resistant to *L. monocytogenes* infection than wild-type mice receiving wild-type bone marrow (Fig. 2e–g). Notably, the wild-type mice transplanted with *Nlrp6*$^{-/-}$ bone marrow and the *Nlrp6*$^{-/-}$ mice that received wild-type bone marrow showed an intermediate level of protection against *L. monocytogenes* (Fig. 2e–g). Indeed, weight loss (Fig. 2e) and bacterial dissemination to the spleen and liver (Fig. 2f, g) of these two chimaera groups were in between the levels seen for the two control groups (Fig. 2e–g). These results suggest that both haematopoietic and

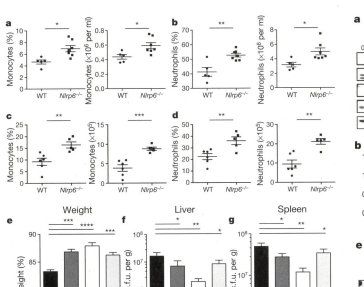


Figure 2 | Enhanced monocyte and neutrophil recruitment in infected *Nlrp6*$^{-/-}$ mice. **a**, **b**, Wild-type and *Nlrp6*$^{-/-}$ mice were infected i.p. with *L. monocytogenes*. Blood was analysed for different cell populations based on their morphology. **c**, **d**, At 6 h after infection, peritoneal lavage was performed on mice infected as above, and the cells collected were analysed as CD11b^{+} Gr1^{-} (also known as Itgam^{+} Ly6g^{-}; monocytes) and CD11b^{+} Gr1^{+} (neutrophils) cells. **e**, Bone marrow chimaera mice were generated for *Nlrp6* as described in Methods. The percentage of weight loss in four groups of chimaeric mice is shown. **f**, **g**, Different groups of chimaeric mice were infected with *L. monocytogenes* and bacterial loads were determined in the liver (**f**) and spleen (**g**) on day 3. Data show mean ± s.e.m. of a representative experiment. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

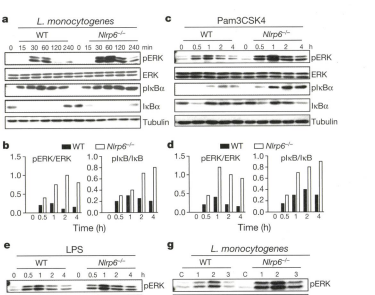


Figure 3 | NLRP6 negatively regulates NF-κB and ERK signalling. **a**, **c**, **e**, Wild-type and *Nlrp6*$^{-/-}$ macrophages were exposed to *L. monocytogenes* or to Pam3CSK4 or LPS, and samples were immunoblotted with the indicated antibodies. **b**, **d**, **f**, Bar graphs showing semi-quantification of phosphorylated ERK (pERK) and pIκB levels relative to ERK and IκB levels, respectively. **g**, Liver lysates from wild-type and *Nlrp6*$^{-/-}$ mice were examined for activation of NF-κB and MAPK signalling by western blot analysis. Each lane represents an individual mouse; C denotes an uninfected mouse. **h**, Bar graphs showing semi-quantification of pERK and pIκB levels relative to ERK and IκB levels, respectively.

non-haematopoietic cells contribute to NLRP6-mediated inhibition of bacterial clearance.

To gain understanding of the signalling mechanisms by which NLRP6 regulates bacterial infections, we first examined bacterial clearance by *in vitro* cultured *Nlrp6*$^{-/-}$ macrophages. Phagocytosis and bacterial replication of *L. monocytogenes* in *Nlrp6*$^{-/-}$ bone marrow-derived macrophages (BMDMs) were comparable to those of wild-type macrophages (Supplementary Fig. 10). Unlike results from overexpression studies[14], but in agreement with a recent report that used *NLRP6* short interfering RNA (siRNA) knockdown in human PBMCs[18], we found NLRP6 to be dispensable for inflammasome activation by *L. monocytogenes* and *S. typhimurium* because neither caspase-1 processing nor maturation of IL-1β was affected in *Nlrp6*-deficient macrophages infected with these pathogens or stimulated with the bacterial ligands lipopolysaccharide (LPS) and the synthetic lipoprotein Pam3CSK4 in the presence of ATP (Supplementary Fig. 11).

We next assessed the potential role of NLRP6 in regulating activation of NF-κB and MAPK signalling in macrophages infected with *L. monocytogenes*. In wild-type BMDMs, the phosphorylation of extracellular signal-regulated kinases 1 and 2 (ERK1/2) and IκB peaked 60 min after infection and gradually diminished thereafter (Fig. 3a, b). Infected *Nlrp6*$^{-/-}$ macrophages had significantly increased phospho-ERK1/2 levels, and failed to downregulate phospho-IκB (Fig. 3a, b). In addition to macrophages, *Nlrp6*$^{-/-}$ neutrophils also responded to *L. monocytogenes* infection with increased phospho-IκB and phospho-ERK1/2 levels (Supplementary Fig. 12a), suggesting that NLRP6 may negatively regulate NF-κB and ERK activation downstream of TLRs. In agreement, the TLR2 ligand Pam3CSK4 also induced increased phospho-ERK1/2 and phospho-IκB levels in *Nlrp6*$^{-/-}$ BMDMs (Fig. 3c, d). Similar observations were made using the TLR4 ligand LPS (Fig. 3e, f). Increased phospho-IκB and phospho-ERK1/2 levels were also noted in liver lysates of infected *Nlrp6*$^{-/-}$ mice (Fig. 3g, h). Unlike with TLR agonists, phospho-IκB and phospho-ERK1/2 levels induced after ligation of the cytosolic receptors NOD1 and NOD2 were similar in wild-type and *Nlrp6*$^{-/-}$ cells (Supplementary Fig. 12b, c), suggesting that NLRP6 is specifically involved in negatively regulating TLR-induced NF-κB and MAPK activation. Thus, in marked contrast to overexpression studies[14], our results in *Nlrp6*$^{-/-}$ cells suggest that NLRP6 may diminish rather than enhance TLR-induced activation of NF-κB and ERK signalling. To confirm the role of NLRP6 in regulating these pathways further, we measured the production of NF-κB- and MAPK-dependent cytokine and chemokine levels at different time points. Abundance of tumour necrosis factor (TNF)-α, interleukin (IL)-6 and keratinocyte chemoattractant (KC) transcripts in cell lysates and secreted cytokines in culture supernatants were significantly higher in *Nlrp6*$^{-/-}$ macrophages infected with *L. monocytogenes*, or stimulated with Pam3CSK4 and LPS (Supplementary Fig. 13a–d). Interferon-β transcript levels nearly doubled in LPS-stimulated *Nlrp6*$^{-/-}$ macrophages after 2 h stimulation, but subsequently returned to levels of wild-type cells (Supplementary Fig. 14). Cytotoxicity was limited to baseline levels in both wild-type and *Nlrp6*$^{-/-}$ macrophages, ruling out an important role for cell death induction in differential secretion of cytokines and chemokines (Supplementary Fig. 15). Increased levels

of IL-6 and KC were also measured in serum and the peritoneal cavity of infected *Nlrp6*$^{-/-}$ mice (Supplementary Fig. 13e, f). These results suggest that NLRP6 specifically suppresses TLR-induced NF-κB and MAPK activation during *L. monocytogenes* infection *in vitro* and *in vivo*.

NF-κB belongs to a family of nuclear transcriptional regulators that initiate pro-inflammatory gene expression[19]. To determine whether canonical or non-canonical NF-κB activation is altered in *Nlrp6*-deficient cells, we analysed the phosphorylation status of the non-canonical NF-κB effector p100, whereas p105 phosphorylation status was monitored as a parameter of canonical NF-κB signalling. *L. monocytogenes* infection induced significantly higher levels of phosphorylated p105 in *Nlrp6*$^{-/-}$ macrophages, whereas those of phosphorylated p100 were similar in wild-type and *Nlrp6*-deficient macrophages (Fig. 4a). Similar results were obtained when analysing the isolated cytosolic fraction of *L. monocytogenes*-infected cells (Fig. 4a). Furthermore, this correlated with increased translocation of the NF-κB subunit RelA (also known as p65) in the nuclear fraction of infected *Nlrp6*-deficient cells, whereas the levels of the non-canonical effector RelB were similar in wild-type and *Nlrp6*$^{-/-}$ macrophages (Fig. 4a). Confocal immunofluorescence microscopy of *L. monocytogenes*-infected wild-type and *Nlrp6*$^{-/-}$ cells confirmed the increased nuclear translocation of RelA in the latter genotype (Fig. 4b, c). Increased p105 phosphorylation was also evident in total and cytosolic lysates of Pam3CSK4-stimulated *Nlrp6*$^{-/-}$ macrophages (Fig. 4d). In addition, enhanced RelA translocation was confirmed by western blotting in nuclear lysates and by confocal immunofluorescence staining (Fig. 4d–f). qPCR analysis of interleukin-1 receptor-associated kinase M (IRAK-M) and the NF-κB negative regulator A20 failed to demonstrate differential expression of these negative regulators of TLR signalling in LPS- and Pam3CSK4-stimulated *Nlrp6*-deficient cells (Supplementary Fig. 16). Although further analysis is required to determine the precise level at which NLRP6 negatively regulates TLR-induced responses, our results indicate that NLRP6 suppresses TLR-induced MAPK and canonical NF-κB signalling to dampen the production of pro-inflammatory cytokines and chemokines during bacterial infection. NLRP6 activation consequently leads to increased susceptibility to both intracellular (*L. monocytogenes* and *S. typhimurium*) and extracellular (*E. coli*) as well as Gram-positive (*L. monocytogenes*) and -negative (*E. coli* and *S. typhimurium*) bacterial pathogens. Our bone marrow chimaera experiments suggested that the haematopoietic and non-haematopoietic compartments contribute to the increased resistance of *Nlrp6*$^{-/-}$ mice to infection. We further demonstrated here that NLRP6 has a crucial role in regulating host and immune responses in cells of the myeloid lineage, and further analysis is required to determine the specific non-haematopoietic cell type(s) contributing to NLRP6-mediated host defence signalling *in vivo*.

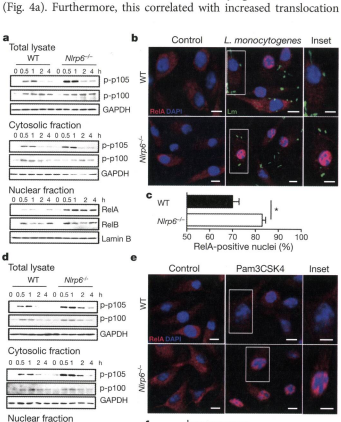


Figure 4 | NLRP6 negatively regulates canonical NF-κB activation pathway. **a**, BMDMs from wild-type and *Nlrp6*$^{-/-}$ mice were infected with *L. monocytogenes* and the phosphorylation of p105 and p100 was examined in total cell lysate and the cytosolic fraction. The nuclear fraction was examined for RelA and RelB activation in wild-type and *Nlrp6*$^{-/-}$ cells. **b**, Cells grown on coverslips were infected with green fluorescent protein (GFP)-labelled *L. monocytogenes* (Lm) and examined for nuclear translocation of RelA using an anti-RelA antibody (red). DAPI, 4′,6-diamidino-2-phenylindole. Scale bars, 10 μm. **c**, Quantitative analysis of RelA-positive nuclei. **d**, Wild-type and *Nlrp6*$^{-/-}$ macrophages were stimulated with Pam3CSK4, and the phosphorylation of p105 and p100 was examined in total cell lysate and the cytosolic fraction. The nuclear fraction was examined for RelA and RelB activation in wild-type and *Nlrp6*$^{-/-}$ cells. **e**, Cells grown on coverslips were stimulated with Pam3CSK4 and examined for nuclear translocation of RelA using an anti-RelA antibody (red). Scale bars, 10 μm. **f**, Quantitative analysis of RelA-positive nuclei. Data show mean ± s.e.m. from three different experiments. *$P < 0.05$.

The detrimental role of NLRP6 during systemic infection with bacterial pathogens appears in sharp contrast to its protective role in gastrointestinal tract, where it contributes to maintaining a healthy composition of the gut microbiota[20]. We propose that these apparently contradictory roles of NLRP6 may be explained by the spatiotemporal context in which it operates. As NLRP6 seems to be essential for dampening the production of pro-inflammatory cytokines and chemokines downstream of TLRs, it may have a protective role under conditions in which strong inflammatory responses might be detrimental to the host (as during colitis). On the other hand, NLRP6 may play an adverse role during situations such as bacterial infections in which potent inflammatory responses would be protective to the host.

Notably, previous studies characterizing the NLR family members NOD1, NOD2, NLRC4 and NLRP3 all pointed to a protective role for these NLRs in clearing microbial pathogens, and deficiency in these NLRs led to increased morbidity and mortality in infected animals[21–24]. By contrast, our data suggest that NLRP6 dampens inflammatory signalling, thereby promoting bacterial dissemination and colonization of systemic organs of the host. In light of these findings, we propose that one subclass of NLRs (including NOD1, NOD2 and NLRC4) may directly respond to microbial pathogens detected in cytosolic compartment by initiating pro-inflammatory signalling that eventually contributes to bacterial clearance. A second subclass of NLR proteins, such as NLRP6, may act as a molecular switch to dampen or silence TLR-induced pathways triggered by extracellular recognition of bacterial ligands to prevent overt pathology. Regardless, our results highlight a crucial role for NLRP6 in dampening host responses against several bacterial pathogens and suggest that therapeutic inhibition of NLRP6 activation may prove beneficial for treating infectious diseases.

METHODS SUMMARY

Nlrp6 knockout mice were generated as described previously[2–4,13]. *In vivo* infection experiments were done on female mice that were 8–12 weeks old and were

conducted under protocols approved by the St. Jude Children's Research Hospital Committee on Use and Care of Animals. Lethal *L. monocytogenes* infection was established by infecting female mice with 1×10^6 c.f.u. bacteria administered i.p. To study bacterial clearance, mice were infected with *L. monocytogenes* at a dose of 3×10^5 c.f.u. (i.p.). Immune cell infiltration in the circulation of infected mice was performed on blood collected by submandibular venipuncture and analysed on a Forcyte haematology analyser. Bone marrow chimaera experiments were done as described previously[25]. Bone marrow macrophages were prepared and infected as described previously[7]. Samples were run on SDS–PAGE and immunoblotted with antibodies as reported previously[7]. Data are expressed as mean ± s.e.m. Differences were analysed by a Student's *t*-test. *P* values ≤ 0.05 were considered significant.

Full Methods and any associated references are available in the online version of the paper.

Received 29 November 2011; accepted 23 May 2012.
Published online 1 July 2012.

1. Kanneganti, T. D., Lamkanfi, M. & Nunez, G. Intracellular NOD-like receptors in host defense and disease. *Immunity* **27,** 549–559 (2007).
2. Chen, G. Y., Liu, M., Wang, F., Bertin, J. & Nunez, G. A functional role for Nlrp6 in intestinal inflammation and tumorigenesis. *J. Immunol.* **186,** 7187–7194 (2011).
3. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145,** 745–757 (2011).
4. Normand, S. *et al.* Nod-like receptor pyrin domain-containing protein 6 (NLRP6) controls epithelial self-renewal and colorectal carcinogenesis upon injury. *Proc. Natl Acad. Sci. USA* **108,** 9601–9606 (2011).
5. Centers for Disease Control and Prevention. CDC Estimates of Foodborne Illness in the United States, 2011 Estimates. http://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html.
6. Scallan, E. *et al.* Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Dis.* **17,** 7–15 (2011).
7. Anand, P. K. *et al.* TLR2 and RIP2 pathways mediate autophagy of *Listeria monocytogenes* via ERK activation. *J. Biol. Chem.* **286,** 42981–42991 (2011).
8. Corr, S. C. & O'Neill, L. A. *Listeria monocytogenes* infection in the face of innate immunity. *Cell. Microbiol.* **11,** 703–709 (2009).
9. Hybiske, K. & Stephens, R. S. Exit strategies of intracellular pathogens. *Nature Rev. Microbiol.* **6,** 99–110 (2008).
10. Creagh, E. M. & O'Neill, L. A. TLRs, NLRs and RLRs: a trinity of pathogen sensors that co-operate in innate immunity. *Trends Immunol.* **27,** 352–357 (2006).
11. Kanneganti, T. D. Central roles of NLRs and inflammasomes in viral infection. *Nature Rev. Immunol.* **10,** 688–698 (2010).
12. Pétrilli, V., Dostert, C., Muruve, D. A. & Tschopp, J. The inflammasome: a danger sensing complex triggering innate immunity. *Curr. Opin. Immunol.* **19,** 615–622 (2007).
13. Henao-Mejia, J. *et al.* Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482,** 179–185 (2012).
14. Grenier, J. M. *et al.* Functional screening of five PYPAF family members identifies PYPAF5 as a novel regulator of NF-κB and caspase-1. *FEBS Lett.* **530,** 73–78 (2002).
15. Stevenson, M. M., Kongshavn, P. A. & Skamene, E. Genetic linkage of resistance to *Listeria monocytogenes* with macrophage inflammatory responses. *J. Immunol.* **127,** 402–407 (1981).
16. Lara-Tejero, M. *et al.* Role of the caspase-1 inflammasome in *Salmonella typhimurium* pathogenesis. *J. Exp. Med.* **203,** 1407–1412 (2006).
17. Serbina, N. V. & Pamer, E. G. Monocyte emigration from bone marrow during bacterial infection requires signals mediated by chemokine receptor CCR2. *Nature Immunol.* **7,** 311–317 (2006).
18. Meixenberger, K. *et al. Listeria monocytogenes*-infected human peripheral blood mononuclear cells produce IL-1β, depending on listeriolysin O and NLRP3. *J. Immunol.* **184,** 922–930 (2010).
19. Waterfield, M. R., Zhang, M., Norman, L. P. & Sun, S. C. NF-κB1/p105 regulates lipopolysaccharide-stimulated MAP kinase signaling by governing the stability and function of the Tpl2 kinase. *Mol. Cell* **11,** 685–694 (2003).
20. van Ooij, C. Immunology: NLRP6 keeps the bad bacteria at bay. *Nature Rev. Microbiol.* **9,** 481 (2011).
21. Girardin, S. E. *et al.* Nod1 detects a unique muropeptide from Gram-negative bacterial peptidoglycan. *Science* **300,** 1584–1587 (2003).
22. Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307,** 731–734 (2005).
23. Franchi, L. *et al.* Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1β in salmonella-infected macrophages. *Nature Immunol.* **7,** 576–582 (2006).
24. Broz, P. *et al.* Redundant roles for inflammasome receptors NLRP3 and NLRC4 in host defense against *Salmonella*. *J. Exp. Med.* **207,** 1745–1755 (2010).
25. Zaki, M. H. *et al.* The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32,** 379–391 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Coyle, E. Grant and R. Flavell for the supply of mutant mice. M.L. is supported by grants from the European Union Framework Program 7 (Marie-Curie Grant 256432), European Research Council (grant 281600) and the Fund for Scientific Research Flanders (G030212N, 1.2.201.10.N.00 and 1.5.122.11.N.00). This work was supported by National Institute of Health Grants (AR056296 and AI101935), and the American Lebanese Syrian Associated Charities (ALSAC) to T.-D.K.

Author Contributions T.-D.K., M.L. and P.K.A. designed the research; P.K.A., R.K.S.M. and J.R.L. performed the experiments; P.V. performed and analysed the histopathology data; J.B. provided essential reagents; T.-D.K., M.L., P.K.A., R.K.S.M., J.R.L., P.V. and J.B. analysed the data; P.K.A., M.L. and T.-D.K. wrote the paper; and T.-D.K. conceived the study, designed the experiments and provided overall direction. M.L. and T.-D.K. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.-D.K. (thirumala-devi.kanneganti@stjude.org).

LETTER

doi:10.1038/nature11263

APJ acts as a dual receptor in cardiac hypertrophy

Maria Cecilia Scimia[1]*, Cecilia Hurtado[1]*, Saugata Ray[1], Scott Metzler[2], Ke Wei[1], Jianming Wang[1], Chris E. Woods[3], Nicole H. Purcell[4], Daniele Catalucci[5,6], Takeshi Akasaka[1], Orlando F. Bueno[7], George P. Vlasuk[7]†, Perla Kaliman[8], Rolf Bodmer[1], Layton H. Smith[1], Euan Ashley[3], Mark Mercola[1], Joan Heller Brown[4] & Pilar Ruiz-Lozano[1,2]

Cardiac hypertrophy is initiated as an adaptive response to sustained overload but progresses pathologically as heart failure ensues[1]. Here we report that genetic loss of APJ, a G-protein-coupled receptor, confers resistance to chronic pressure overload by markedly reducing myocardial hypertrophy and heart failure. In contrast, mice lacking apelin (the endogenous APJ ligand) remain sensitive, suggesting an apelin-independent function of APJ. Freshly isolated APJ-null cardiomyocytes exhibit an attenuated response to stretch, indicating that APJ is a mechanosensor. Activation of APJ by stretch increases cardiomyocyte cell size and induces molecular markers of hypertrophy. Whereas apelin stimulates APJ to activate $G\alpha_i$ and elicits a protective response, stretch signals in an APJ-dependent, G-protein-independent fashion to induce hypertrophy. Stretch-mediated hypertrophy is prevented by knockdown of β-arrestins or by pharmacological doses of apelin acting through $G\alpha_i$. Taken together, our data indicate that APJ is a bifunctional receptor for both mechanical stretch and the endogenous peptide apelin. By sensing the balance between these stimuli, APJ occupies a pivotal point linking sustained overload to cardiomyocyte hypertrophy.

G-protein-coupled receptors (GPCRs) have been widely implicated in the control of cardiac function. These receptors couple to heterotrimeric GTP-binding proteins of the $G\alpha_s$, $G\alpha_i$, $G\alpha_{q/11}$ and $G\alpha_{12/13}$ families, and transduce the GPCR signal to intracellular targets. Numerous studies have linked $G\alpha_s$ to increased contractility, $G\alpha_{q/11}$ to pathological hypertrophy[2,3] and $G\alpha_i$ to cardioprotection[4]. APJ is a GPCR identified as the receptor for the adipokine apelin[5,6]. Apelin-activated APJ signals through $G\alpha_i$, exerting a positive effect on cardiac contractility[7–9] and a vasodilator activity that counteracts angiotensin-II-induced atheroma[10,11]. Apelin administration blunts progression to hypertrophy (Supplementary Fig. 1 and Supplementary Tables 2 and 3) and apelin-KO mice show susceptibility to heart failure[12] (also see Supplementary Fig. 1 and Supplementary Table 1). Thus, apelin and its receptor APJ are emerging as potential therapeutic targets.

We examined the response of APJ knockout (APJ-KO) mice to sustained pressure overload by transaortic constriction (TAC). Although deletion of APJ resulted in some prenatal lethality[13,14], all viable APJ-KO mice showed normal adult appearance and cardiovascular parameters at baseline (Supplementary Table 4). However, APJ-null animals were resistant to the pathological hypertrophic response to TAC (Fig. 1a–d) observed both in wild type (WT) and in apelin-KO mice (Supplementary Fig. 1g–i). APJ-KO mice responded to TAC by initially increasing cardiac mass, but the maladaptive progression to dilated ventricular hypertrophy was blunted shortly after injury (Supplementary Table 4). The protective effect persisted long-term (Fig. 1a, b, g, h) in all parameters measured, including diminished cardiomyocyte size (Fig. 1c, d), reduced fibrosis (Fig. 1e, f), sustained cardiac contractility (Fig. 1g) relative to WT and apelin-KO mice (Supplementary Tables 1 and 4) and reduced heart weight/body weight ratio (Fig. 1h). Baseline cardiac contractility, measured as percentage fractional shortening, was approximately 38% across genotypes. After 90 days of TAC, percentage fractional shortening decreased to 22 ± 2% in WT, 23 ± 1% in apelin KO mice, but remained at 34 ± 2% in APJ-KO mice ($P = 0.01$ between APJ-KO

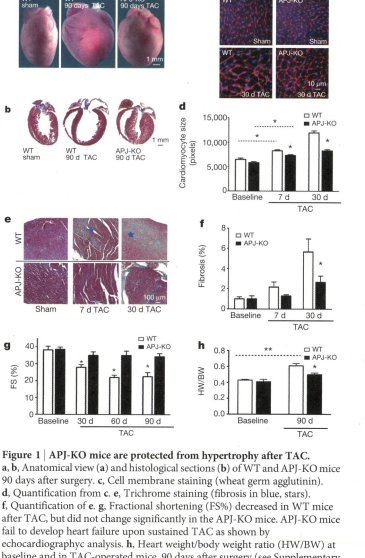


Figure 1 | APJ-KO mice are protected from hypertrophy after TAC. **a**, **b**, Anatomical view (**a**) and histological sections (**b**) of WT and APJ-KO mice 90 days after surgery. **c**, Cell membrane staining (wheat germ agglutinin). **d**, Quantification from **c**. **e**, Trichrome staining (fibrosis in blue, stars). **f**, Quantification of **e**. **g**, Fractional shortening (FS%) decreased in WT mice after TAC, but did not change significantly in the APJ-KO mice. APJ-KO mice fail to develop heart failure upon sustained TAC as shown by echocardiographyc analysis. **h**, Heart weight/body weight ratio (HW/BW) at baseline and in TAC-operated mice, 90 days after surgery (see Supplementary Table 4 for details). Error bars, s.e.m.*$P < 0.05$ between indicated groups, analysis of variance.

[1]Sanford-Burnham Medical Research Institute, La Jolla, California 92037, USA. [2]Department of Pediatrics, School of Medicine, Stanford University, California 94305, USA. [3]Department of Medicine, School of Medicine, Stanford University, California, 94305, USA. [4]Department of Pharmacology, University of California, San Diego, California, USA. [5]Biomedical and Genetic Research Institute, National Research Council, via Fantoli 16/15, 20138, Milan, Italy. [6]Humanitas Clinical and Research Center, Via Manzoni 56, 20089 Rozzano, Milan, Italy. †Wyeth Pharmaceutical, Madison, New Jersey, USA. [8]Institute of Biomedical Research, August Pi i Sunyer (IDIBAPS), E-08036 Barcelona, Spain. †Present address: Sirtris, a GSK Company, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

and WT) (Fig. 1g and Supplementary Tables 1 and 4). In summary, both WT and apelin-KO mice presented clear signs of heart failure after 90 days of TAC, whereas APJ-KO mice were almost unaffected. The maintenance of cardiac function in the APJ-KO demonstrates that the expression of APJ is necessary to elicit heart failure in response to pressure overload.

The different responses of apelin-KO and APJ-KO mice to TAC imply that either apelin can act independently of APJ, or that APJ transduces a signal independently of apelin. We tested the first hypothesis by infusing APJ-KO mice with apelin (285 $\mu g\,kg^{-1}$ per 24 h) and examining two readouts: contractility under TAC and vascular tone. Notably, apelin infusion did not increase cardiac contractility (percentage fractional shortening) in TAC-APJ-KO mice, in contrast to the characteristic improvement seen in TAC-WT animals (Supplementary Fig. 2a). In the absence of apelin infusion, endogenous levels of apelin in blood increased after TAC from 1 to 2 $ng\,ml^{-1}$ and that rise was not different in WT and APJ-KO mice, making it unlikely that the protection achieved in the APJ-KO is due to hyper-activation of apelin signalling (Supplementary Fig. 2b). To test vascular tone, systolic and diastolic blood pressures were increased by infusion of Ang·II (1,000 $ng\,kg^{-1}\,min^{-1}$). Apelin infusion significantly decreased systolic blood pressure in WT animals but not in APJ-KO mice (Supplementary Fig. 2c–f), further indicating that apelin activity requires APJ.

Because the mechanical properties of the heart change markedly during pressure overload[15], and the structurally related angiotensin receptor (AT-1) can act as a mechanosensor[16], we asked whether APJ responds to mechanical stretch. Initially these experiments were challenging as cultured cardiomyocytes consistently downregulate the expression of endogenous APJ (Supplementary Fig. 3a), and studies had to be restricted to freshly isolated adult cardiomyocytes. We mimicked the effect of pressure overload by using a carbon fibre technique[17] to stretch cardiomyocytes and evaluated their Frank–Starling gain (FSG). FSG is a dimensionless metric of the force that can be recruited by stretch[18]. Freshly isolated adult cardiomyocytes from WT mice showed a significantly higher FSG than cardiomyocytes from APJ-KO mice (Fig. 2a–c). Treatment with apelin decreased the FSG in WT cardiomyocytes but showed no effect in APJ-KO cells (Fig. 2d). Therefore, apelin modulated the response to stretch only in cardiomyocytes with APJ receptors.

Engineered cells stably expressing human APJ (APJ-HEK) responded to apelin by increasing the content of phosphorylated extracellular signal-regulated kinase (pERK) (Fig. 3a, b) whereas parental (HEK) cells showed no significant change. Stretch also increased pERK content in cells expressing APJ (Fig. 3a, b). pERK levels, therefore, reflect the cellular response to APJ activation by either stretch or apelin. pERK was therefore used as a simple readout of APJ activation. The $G\alpha_i$ inhibitor PTX blocked the ability of apelin, but not stretch, to induce APJ-dependent phosphorylation of ERK (Fig. 3a, b), first suggesting that separate mechanisms link stretch and apelin to APJ intracellular signalling.

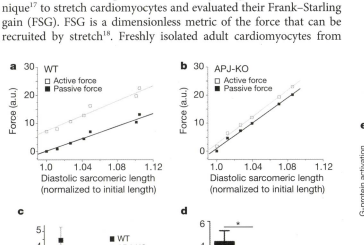


Figure 2 | APJ mediates a stretch response that can be modulated by apelin. **a**, **b**, Representative force measurements (arbitrary units) for end-diastolic and end-systolic length–tension relationships in adult cardiomyocytes from WT (**a**) and APJ-KO (**b**) mice plotted against diastolic length (normalized to unstretched length). Cells were paced at 1 Hz. **c**, FSG attained by dividing the active force by the passive force from experiments in **a** and **b** plotted as a function of diastolic length ($n = 8$ WT and 7 APJ-KO cardiomyocytes). **d**, Average FSG (at 1.02 sarcomeric length) from **c** is shown for APJ-KO and WT both with (+) and without (−) 10 nM apelin administration ($n = 6$ for WT + apelin and $n = 7$ for APJ-KO + apelin). Error bars, s.e.m.

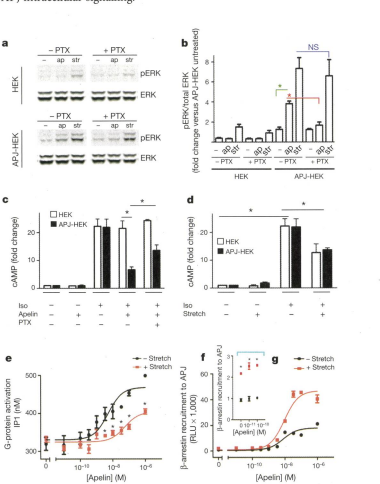


Figure 3 | Stretch activation of APJ enhances β-arrestin while reducing G-protein signalling. **a**, **b**, Immunoblot (**a**) and quantification (**b**) of ERK from APJ stably transfected (APJ-HEK) and parental (HEK) cells treated for 5 min with 100 nM apelin or stretch in absence or presence of PTX ($G\alpha_i$ inhibitor), $n = 3$. **c**, **d**, Effect of 1 μM apelin (**c**) or stretch (**d**) on cAMP levels. Isoproterenol (iso, 1 μM) was used to elevate cAMP artificially to study $G\alpha_i$ activation, $n = 4$. **e**, G-protein ($G\alpha_{q,s,i}$ and $G\alpha_{12/13}$) activation by apelin and effect of stretch (red) in CHO cells expressing APJ and $G\alpha_{16}$. Receptor stimulation activated the promiscuous $G\alpha_{16}$, phospholipase-Cβ and caused the accumulation of IP1 (representative of three experiments, $n = 4$ samples). **f**, **g**, Arrestin recruitment to the APJ receptor in response to apelin (black) and stretch (red) by an enzyme complementation assay in CHO cells expressing recombinant APJ and β-arrestin2 (representative of three experiments, $n = 3$ samples). **f**, Full range of β-arrestin binding to APJ under either physiological or pharmacological doses of apelin. **g**, Data points for $0–10^{-10}$ M apelin. RLU, relative light units. Error bars, s.e.m. $^*P < 0.05$ between indicated groups, analysis of variance.

We next examined whether activation of APJ by stretch or apelin differentially modulates G-protein-generated second messengers. Activation of $G\alpha_i$ is associated with inhibition of adenyl-cyclase and reduced cAMP, whereas activation of $G\alpha_s$ increases cAMP[19]. cAMP did not increase after application of stretch or apelin, arguing against activation of adenyl-cyclase (and $G\alpha_s$) (Fig. 3c, d, no isoproterenol conditions). In contrast, when isoproterenol was used to elevate the intracellular concentration of cAMP[20], apelin addition decreased cAMP levels in the APJ-HEK cells (Fig. 3c). This effect of apelin was partly inhibited by PTX, consistent with the involvement of $G\alpha_i$ (Fig. 3c). Apelin did not modify isoproterenol-stimulated cAMP formation in untransfected HEK controls, showing that the decrease is mediated by APJ. These data agree with previous reports[21] and demonstrate that $G\alpha_i$ transduces the signal initiated by apelin binding to APJ. In contrast, stretch reduced cAMP in parental HEK as well as in APJ-HEK cells (Fig. 3d), indicating that, although stretch can activate $G\alpha_i$ signalling, this response is not mediated through APJ.

The $G\alpha_{16}$ subunit couples any activated GPCR to phospholipase C, which results in accumulation of inositol phosphates (IP1), and thus provides a general readout of G-protein activation[22]. Stretch did not

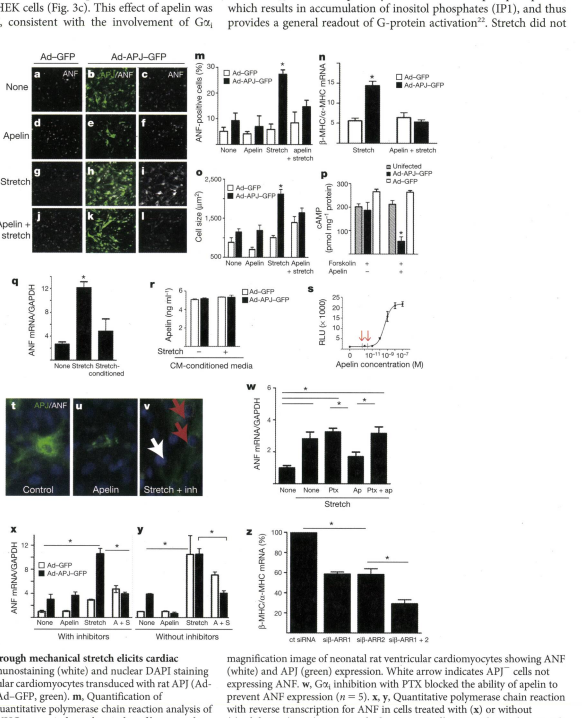


Figure 4 | APJ activation through mechanical stretch elicits cardiac hypertrophy. **a–l**, ANF immunostaining (white) and nuclear DAPI staining (blue) of rat neonatal ventricular cardiomyocytes transduced with rat APJ (Ad-APJ–GFP) or control GFP (Ad–GFP, green). **m**, Quantification of **a–l** ($n = 250–350$ cells). **n**, Quantitative polymerase chain reaction analysis of the ratio between β- and α-MHCs, as an independent index of hypertrophy ($n = 3–5$ samples). **o**, Mean cell sizes in **a–l** ($n = 24$ cells). **p**, Responsiveness of transfected cells to apelin treatment ($n = 4–5$). **q**, ANF expression in cardiomyocytes in the presence of conditioned media from stretched cardiomyocytes ($n = 3$). **r**, Apelin enzyme-linked immunosorbent assay of conditioned media (12 h) from cardiomyocytes (CM) non-stretched and stretched. **s**, Apelin standard curve of CHO APJ–β-arrestin interaction assay; red arrows represent the response elicited by two samples from **r**. **t–v**, Higher magnification image of neonatal rat ventricular cardiomyocytes showing ANF (white) and APJ (green) expression. White arrow indicates APJ$^-$ cells not expressing ANF. **w**, $G\alpha_i$ inhibition with PTX blocked the ability of apelin to prevent ANF expression ($n = 5$). **x**, **y**, Quantitative polymerase chain reaction with reverse transcription for ANF in cells treated with (**x**) or without (**y**) inhibitors ($n = 3$). **z**, Diminished expression of hypertrophy markers in Ad-APJ–GFP cardiomyocytes transfected upon knockdown of β-arrestin1 (siβ-ARR1), β-arrestin2 (siβ-ARR2) or both (siβ-ARR1+2) ($n = 4$). Except for **p** and **y**, all experiments were performed in the presence of inhibitors of AT-1 (candesartan, 100 nM) and ET-A (BQ123, 300 nM) added 1 h before stretch and/or apelin treatment until fixation. All are representative experiments performed at least three independent times. Error bars, s.e.m. $^*P < 0.05$ between indicated groups, analysis of variance.

increase IP1 production whereas apelin did so in a dose-responsive manner (Fig. 3e). Notably, for any given concentration of apelin, stretch consistently reduced G-protein activation (Fig. 3e). There was a decrease in maximal levels and a shift in half-maximum effective concentration (EC_{50}) from 5.1×10^{-9} to 5.5×10^{-8} when both stimuli were applied simultaneously (Fig. 3e). These experiments demonstrate that APJ activation by stretch is largely G-protein independent and that stretch interferes with apelin-mediated G-protein activation by APJ.

Using a β-arrestin/APJ complementation assay, apelin was found to induce a dose-dependent increase in β-arrestin signalling, as expected (Fig. 3f). Stretch in the absence of apelin boosted β-arrestin signalling twofold more than baseline (Fig. 3g). Notably, stretch increased β-arrestin recruitment in response to apelin at all doses tested (Fig. 3g). Taken together, stretch causes APJ to signal less effectively/potently through G proteins, but to recruit β-arrestin more effectively.

To investigate whether these distinct mechanisms of APJ action differentially affect cardiac hypertrophy, we examined cardiomyocyte cell growth and the expression of molecular markers of pathological hypertrophy in neonatal rat ventricular cardiomyocytes[23,24]. These cardiomyocytes in culture respond to mechanical stretch[25], mainly through the angiotensin II receptor AT-1[26] and the endothelin1 receptor ET-A[27], which are also GPCRs. To test the specific effect of APJ in cardiac stretch/hypertrophy, we used pharmacological inhibitors of AT-1 (100 nM candesartan) and ET-A (300 nM BQ123), hereafter labelled as 'inhibitors'. To overcome APJ down-regulation during cardiomyocyte culture, we re-established APJ expression by adenoviral transduction (about 90% efficiency) with either control GFP (Ad–GFP) or an APJ–GFP fusion protein (Ad-APJ–GFP). Hypertrophy was assessed by the characteristic increase in perinuclear immunolocalization of atrial natriuretic factor (ANF) (Fig. 4a–m). Remarkably, apelin treatment did not increase ANF immunostaining, neither in Ad-APJ–GFP- nor in Ad–GFP-infected cells (Fig. 4d–f, m). In contrast, stretch applied in the presence of inhibitors significantly increased the number of perinuclear ANF^+ cells in APJ-restored cardiomyocytes (27 ± 1%), but not in control cardiomyocytes infected with Ad–GFP (5.8 ± 2.1%) (Fig. 4g–i, m). Co-stimulation with apelin and stretch reduced the number of ANF^+ cells (Fig. 4j–l, m). Stretch also induced specific changes in the expression of other molecular markers of hypertrophy, including an increased ratio of β- to α-myosin heavy chain (β-MHC/α-MHC) (Fig. 4n) and increased cell size (Fig. 4o). Similar to its effects on ANF, the addition of apelin also attenuated effects of stretch on other parameters of hypertrophy (Fig. 4m–o). The induction of ANF by stretch was PTX insensitive, but the ability of apelin to antagonize stretch-induced hypertrophy was prevented by treatment with PTX (Fig. 4w).

Several control experiments confimed that APJ is directly involved in the response to stretch: neonatal rat ventricular cardiomyocytes treated with forskolin responded to apelin only upon reconstitution of APJ expression, as monitored by their ability to decrease intracellular cAMP levels (Fig. 4p). The possibility that stretch induces the secretion of factors that might indirectly activate hypertrophy through APJ was considered but seems unlikely because conditioned medium from APJ-transfected cardiomyocytes did not induce ANF expression (Fig. 4q). Moreover, the concentration of apelin in the media remained unchanged (approximately 5 ng ml^{-1}) with or without stretch (Fig. 4r), and conditioned media from stretched cells could not activate β-arrestin recruitment (Fig. 4s). Single-cell analysis of low multiplicity APJ–GFP transduced cells showed that cardiomyocytes require APJ to induce ANF upon stretch (Fig. 4t–v, note perinuclear ANF immunostaining only in GFP^+ cells in panel v), further demonstrating that APJ activation through stretch is sufficient to elicit cardiac hypertrophy. Previous studies showed that APJ interacts with AT-1 and apelin antagonizes AT-1 function[11]. As shown in Fig. 4x, in the presence of AT-1 inhibitors, only those cells expressing APJ responded to stretch by significantly increasing perinuclear ANF expression. In the absence of AT-1 inhibitors, APJ-transduced and non-transduced cells reached the same maximal level of ANF expression (Fig. 4y). Thus, blocking AT-1 does not impair the ability of APJ to respond to stretch, indicating that APJ alone is sufficient to transduce a stretch-induced hypertrophy signal.

Notably, siRNA specific for β-arrestin1 or β-arrestin2 (Fig. 4z) blocked the stretch induction of hypertrophic markers (β-MHC/α-MHC (Fig. 4z) and ANF (not shown)) with an additive effect when both siRNAs were used together. These data substantiate the model that APJ signalling through β-arrestin mediates stretch-induced myocardial hypertrophy.

In summary, the mechano-response of APJ is necessary (blunted hypertrophic response to TAC of APJ-KO mice, Fig. 1) and sufficient (stretch induction of ANF expression occurs in cells expressing APJ, Fig. 4) to trigger myocardial hypertrophy in a β-arrestin-dependent manner (Fig. 4z). Apelin does not induce hypertrophy, but instead blunts stretch-mediated hypertrophy (Fig. 4j–o), suggesting the ability of apelin to override pathological signalling from stretch. At a mechanistic level, APJ transduces apelin and stretch signalling differently. The response to apelin seems to be G-protein-PTX sensitive, whereas that induced by stretch is PTX insensitive and G-protein independent in the absence of exogenous apelin (Fig. 3). Stretch profoundly affects apelin signalling, diminishing G-protein activation while augmenting β-arrestin recruitment (Fig. 3e–g). These data indicate that APJ integrates apelin and stretch stimuli, biasing the levels of G-protein signalling versus β-arrestin recruitment accordingly.

These results have implications for the consideration of APJ as a drug target, because APJ/stretch can be pathological. Therefore, a beneficial effect will be obtained not by general apelin receptor agonism, but rather by selectively inhibiting the ability of APJ to respond to mechanical stretch or by blocking its interaction with molecules that initiate pathological signalling cascades.

METHODS SUMMARY

All experiments were performed in accordance with relevant guidelines and regulations. The Sanford-Burnham Medical Research Institute's Animal Care & Use Program is accredited by Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC International), and a Multiple Project Assurance A3053-1 is on file in the Office of Laboratory Animal Welfare, Department of Health and Human Services.

APJ and apelin KO mice. APJ-KO mice were obtained from Deltagen. Apelin-KO mice are described elsewhere[13]. Both APJ and apelin lines (male and female) were in C57Bl/6 genetic background in a 99–100% purity, as demonstrated by microsatellite analysis (Radil). For detailed methods, see Supplementary Information.

Received 9 March; accepted 25 May 2012.
Published online 18 July; corrected online 15 August 2012 (see full-text HTML for details).

1. Levy, D., Garrison, R. J., Savage, D. D., Kannel, W. B. & Castelli, W. P. Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. *N. Engl. J. Med.* **322,** 1561–1566 (1990).
2. Knowlton, K. U. *et al.* The alpha 1A-adrenergic receptor subtype mediates biochemical, molecular, and morphologic features of cultured myocardial cell hypertrophy. *J. Biol. Chem.* **268,** 15374–15380 (1993).
3. Rapacciuolo, A., Esposito, G., Prasad, S. V. & Rockman, H. A. G protein-coupled receptor signalling in in vivo cardiac overload. *Acta Physiol. Scand.* **173,** 51–57 (2001).
4. Rockman, H. A., Koch, W. J. & Lefkowitz, R. J. Seven-transmembrane-spanning receptors and heart function. *Nature* **415,** 206–212 (2002).
5. O'Dowd, B. F. *et al.* A human gene that shows identity with the gene encoding the angiotensin receptor is located on chromosome 11. *Gene* **136,** 355–360 (1993).
6. Tatemoto, K. *et al.* Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. *Biochem. Biophys. Res. Commun.* **251,** 471–476 (1998).
7. Szokodi, I. *et al.* Apelin, the novel endogenous ligand of the orphan receptor APJ, regulates cardiac contractility. *Circ. Res.* **91,** 434–440 (2002).
8. Ashley, E. A. *et al.* The endogenous peptide apelin potently improves cardiac contractility and reduces cardiac loading *in vivo*. *Cardiovasc. Res.* **65,** 73–82 (2005).
9. Jia, Y. X. *et al.* Apelin protects myocardial injury induced by isoproterenol in rats. *Regul. Pept.* **133,** 147–154 (2006).
10. Siddiquee, K. *et al.* Apelin protects against angiotensin II-induced cardiovascular fibrosis and decreases plasminogen activator inhibitor type-1 production. *J. Hypertens.* **29,** 724–731 (2011).

11. Chun, H. J. *et al.* Apelin signaling antagonizes Ang II effects in mouse models of atherosclerosis. *J. Clin. Invest.* **118,** 3343–3354 (2008).
12. Kuba, K. *et al.* Impaired heart contractility in apelin gene-deficient mice associated with aging and pressure overload. *Circ. Res.* **101,** e32–e42 (2007).
13. Charo, D. N. *et al.* Endogenous regulation of cardiovascular function by apelin-APJ. *Am. J. Physiol. Heart Circ. Physiol.* **297,** H1904–H1913 (2009).
14. Scott, I. C. *et al.* The G protein-coupled receptor Agtr1b regulates early development of myocardial progenitors. *Dev. Cell* **12,** 403–413 (2007).
15. Frank, D. *et al.* Gene expression pattern in biomechanically stretched cardiomyocytes: evidence for a stretch-specific gene program. *Hypertension* **51,** 309–318 (2008).
16. Zou, Y. *et al.* Mechanical stress activates angiotensin II type 1 receptor without the involvement of angiotensin II. *Nature Cell Biol.* **6,** 499–506 (2004).
17. Iribe, G. *et al.* Axial stretch of rat single ventricular cardiomyocytes causes an acute and transient increase in Ca^{2+} spark rate. *Circ. Res.* **104,** 787–795 (2009).
18. Bollensdorff, C., Lookin, O. & Kohl, P. Assessment of contractility in intact ventricular cardiomyocytes using the dimensionless 'Frank-Starling Gain' index. *Pflügers Arch.* **462,** 39–48 (2011).
19. Civelli, O. GPCR deorphanizations: the novel, the known and the unexpected transmitters. *Trends Pharmacol. Sci.* **26,** 15–19 (2005).
20. Means, C. K., Miyamoto, S., Chun, J. & Brown, J. H. $S1P_1$ receptor localization confers selectivity for G_i-mediated cAMP and contractile responses. *J. Biol. Chem.* **283,** 11954–11963 (2008).
21. Masri, B., Morin, N., Pedebernade, L., Knibiehler, B. & Audigier, Y. The apelin receptor is coupled to G_{i1} or G_{i2} protein and is differentially desensitized by apelin fragments. *J Biol Chem* **281,** 18317–18326 (2006).
22. Kostenis, E. Is $G\alpha_{16}$ the optimal tool for fishing ligands of orphan G-protein-coupled receptors? *Trends Pharmacol. Sci.* **22,** 560–564 (2001).
23. Knowlton, K. U. *et al.* Divergent pathways mediate the induction of ANF transgenes in neonatal and hypertrophic ventricular myocardium. *J. Clin. Invest.* **96,** 1311–1318 (1995).
24. Rockman, H. A. *et al.* Molecular and physiological alterations in murine ventricular dysfunction. *Proc. Natl Acad. Sci. USA* **91,** 2694–2698 (1994).
25. Sadoshima, J. & Izumo, S. Mechanical stretch rapidly activates multiple signal transduction pathways in cardiac myocytes: potential involvement of an autocrine/paracrine mechanism. *EMBO J.* **12,** 1681–1692 (1993).
26. Sadoshima, J., Xu, Y., Slayter, H. S. & Izumo, S. Autocrine release of angiotensin II mediates stretch-induced hypertrophy of cardiac myocytes in vitro. *Cell* **75,** 977–984 (1993).
27. Magga, J., Vuolteenaho, O., Marttila, M. & Ruskoaho, H. Endothelin-1 is involved in stretch-induced early activation of B-type natriuretic peptide gene expression in atrial but not in ventricular myocytes: acute effects of mixed ET(A)/ET(B) and AT1 receptor antagonists *in vivo* and *in vitro*. *Circulation* **96,** 3053–3062 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. We are thankful to N. Ling, S. Zhao and F. Abdel-Wahhab for technical assistance; E. Sergienko for help optimizing arrestin assay; E. Adamson for reading the manuscript; and M. Querol for graphics design. This work was supported by Wyeth Sponsored Research Agreement and National Institutes of Health (NIH) grant R01HL086879 to P.R.L.; NIH grants R37HL059502 and R01HL083463 and the Sanford Children's Center to M.M.; NIH grant R01HL054732, grants from the Ellison Medical Foundation and the Muscular Dystrophy association to R.B., NIH grant (NS05422) and Florida Department of Health grant 06-NIR-09 to L.H.S., and NIH grants R01HL28143, P01 HL085577 to J.H.B. M.C.S. has received support from the California Institute for Regenerative Medicine (clinical fellow), the Italian Ministry of Research and Education and the Italian Society of Cardiology (SIC and Sanofi-Aventis Foundation). C.H. holds an American Heart Association Postdoctoral Award, S.R. is a Sanford Children's Health Research Center fellow. P.K. is supported by SAF2010-15050 Ministerio de Ciencia e Innovación (MICINN) Spain.

Author Contributions M.S.C. and C.H. designed, performed experiments, analysed data and prepared the manuscript. C.E.W., S.R., S.A.M., K.W., S.M., J.W., N.H.P., T.A. and P.K. designed and performed experiments, and analysed data. D.C., G.P.V., R.B., O.F.B., L.H.S., E.A. and J.H.B. designed experiments. M.M. designed experiments and prepared the manuscript. P.R.L. designed and supervised experiments, analysed data and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.R-L. (prlozano@stanford.edu).

LETTER

doi:10.1038/nature11248

'Slings' enable neutrophil rolling at high shear

Prithu Sundd[1], Edgar Gutierrez[2], Ekaterina K. Koltsova[1], Yoshihiro Kuwano[1,3], Satoru Fukuda[4], Maria K. Pospieszalska[1], Alex Groisman[2] & Klaus Ley[1]

Most leukocytes can roll along the walls of venules at low shear stress (1 dyn cm^{-2}), but neutrophils have the ability to roll at tenfold higher shear stress in microvessels *in vivo*[1,2]. The mechanisms involved in this shear-resistant rolling are known to involve cell flattening[3] and pulling of long membrane tethers at the rear[4–6]. Here we show that these long tethers do not retract as postulated[6,7], but instead persist and appear as 'slings' at the front of rolling cells. We demonstrate slings in a model of acute inflammation *in vivo* and on P-selectin *in vitro*, where P-selectin-glycoprotein-ligand-1 (PSGL-1) is found in discrete sticky patches whereas LFA-1 is expressed over the entire length on slings. As neutrophils roll forward, slings wrap around the rolling cells and undergo a step-wise peeling from the P-selectin substrate enabled by the failure of PSGL-1 patches under hydrodynamic forces. The 'step-wise peeling of slings' is distinct from the 'pulling of tethers' reported previously[4–6,8]. Each sling effectively lays out a cell-autonomous adhesive substrate in front of neutrophils rolling at high shear stress during inflammation.

Neutrophil rolling at high shear stress is poorly understood[2]. Some neutrophil adhesion molecules including PSGL-1 are concentrated around the tips of microvilli, where they initiate contact with the vessel wall[9,10]. When the force on a microvillus exceeds a threshold force of ~31 pN, the plasma membrane eventually separates from the cytoskeleton, forming a tether[9,11]. To study the footprints of cells rolling under high shear, we recently developed quantitative dynamic footprinting (qDF)[4] and refined it to allow simultaneous acquisition of two fluorochromes (DqDF)[12]. qDF relies on total internal reflection fluorescence (TIRF) microscopy of fluorescent neutrophils rolling in microfluidic channels on glass substrates coated with P-selectin. We observed that DiI-stained neutrophils isolated from the bone marrow of wild-type mice (Fig. 1a and Supplementary Movie 1) or neutrophils in whole blood of *Lyz2-EGFP* mice[13] (Fig. 1b and Supplementary Movie 2) rolling at 6–10 dyn cm^{-2} frequently formed long processes in the downstream direction. These structures have not been observed previously, and we propose to call them 'slings'. Three-dimensional-reconstructions of the qDF footprint of a rolling neutrophil with a sling in front are shown in Fig. 1c–d (Supplementary Note 1). Scanning electron micrographs (SEMs; Fig. 1e) confirm that neutrophils rolling at 10 dyn cm^{-2} have slings in front and tethers in the rear. As the cells roll forward, the slings are wrapped around the rolling neutrophils, which was revealed both by qDF (Fig. 1f and Supplementary Movie 3) and SEMs (Fig. 1g).

Slings were observed *in vivo* by epifluorescence intravital microscopy of mouse cremaster venules. Slings in front of six different leukocytes (labelled with a monoclonal antibody (mAb) against LFA-1) rolling in cremaster venules of wild-type mice *in vivo* are shown in Fig. 1h–m, Supplementary Fig. 1 and Supplementary Movies 4 and 5.

When a neutrophil is first captured from the flow, an initial sling may take several seconds to form. Slings become more frequent as the rolling progresses at high shear stress (Fig. 2a, b) and their average length increases with shear stress over the range of 6 to 10 dyn cm^{-2} (Fig. 2c; Supplementary Note 2). Because neutrophil rolling at high shear stress produces three to four tethers per cell[4], we propose that slings may form when detached tethers swing around to the front of the rolling cell. This was confirmed by both qDF (Fig. 2d and Supplementary Movie 6) and differential interference contrast (DIC) microscopy (Fig. 2e and Supplementary Movie 7). Because each tether grows over time, and the aspect ratio of the tether anchorage points suggested increasing force (Supplementary Fig. 2), we calculated the tether force based on the nonlinearly decaying spring viscoelastic (NLDs-viscoelastic) model[9]. Although the anchorage points along newly formed slings initially did not show evidence of deformation, this changed as the cell rolled over the sling, which resulted in the rearmost patch to become load-bearing (white arrowhead at $t = 13$ s in Fig. 1a and $t = 5$ s in Fig. 1b). The force acting on a tether or a load-bearing sling reached a maximum of 80–90 pN just before the tether or sling detached (Fig. 2f). If a rolling cell has two loaded slings and two tethers with the inflection points (white arrowhead at $t = 13$ s in Fig. 1a) and anchorage points (white arrowhead in Supplementary Fig. 2a), respectively, at a distance of 10 μm from the cell centre, then the two slings and two tethers can together balance the entire bond torque and more than three quarters of the forward force acting on the rolling cell (Fig. 2g and Supplementary Fig. 3). Thus, tethers and slings are the major structures which slow down neutrophils rolling at high wall shear stress.

qDF experiments with both naive and Th1 CD4 T cells (Supplementary Note 3; Supplementary Figs 4 and 5) revealed that the formation of slings is indeed associated with stabilized rolling at high shear stress. Next, we addressed the nature of PSGL-1 expression on slings. Raw DqDF images of wild-type neutrophils rolling on P-selectin (Supplementary Fig. 6a) show expression of PSGL-1 (4RB12 mAb[14]; green) on the cell surface, sling and tether anchorage points. Consistent with previous results in suspended cells[9,10], PSGL-1 is localized around the tips of microvilli of rolling neutrophils (Supplementary Fig. 6b), but excluded from the cell surface. As anticipated, PSGL-1 was present in the tether anchorage points, but surprisingly, PSGL-1 was also expressed in patches along the sling. Further experiments were conducted using a red fluorochrome (Alexa-Fluor-568) coupled to 4RB12 mAb (Fig. 3a–c and Supplementary Figs 7 and 8). Specificity was established by investigating PSGL-1-deficient neutrophils rolling on E-selectin which showed no staining for PSGL-1 (Supplementary Fig. 9). We observed a high concentration of PSGL-1 not only at the tip (the former tether anchorage point), but also in discrete patches along the length of slings (Fig. 3a and Supplementary Figs 7 and 8). Analysis of SEM images of slings revealed tips of microvilli preserved as lobes on the slings (Supplementary Fig. 10 and 11). These findings indicate that membrane from several microvilli is pulled from the cell surface to form the tethers. Tethers following detachment become slings and the tips of microvilli, which are rich in PSGL-1, appear as patches on the slings. Indeed, the spacing of PSGL-1 patches on slings (1.6 ± 0.2 (s.e.m.) μm; $n = 34$) is comparable to the estimated distance between the tips of adjacent microvilli (1.32 μm) preserved as lobes on a 10 μm long sling (Supplementary Fig. 12).

[1]Division of Inflammation Biology, La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA. [2]Department of Physics, University of California San Diego, La Jolla, California 92093, USA. [3]Department of Dermatology, University of Tokyo, Tokyo 113-8655, Japan. [4]Laboratory of Electron Microscopy, University of Tokyo, Tokyo 113-8655, Japan.

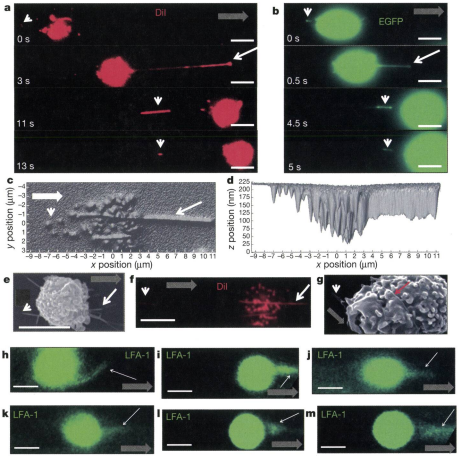


Figure 1 | Rolling neutrophils form slings. a, b, Processed qDF images of a DiI-stained isolated neutrophil (**a**) and an *EGFP* neutrophil in whole blood (**b**) rolling on P-selectin in a microfluidic device. A tether (anchor point marked with white arrowhead at $t = 0$ s) swings around to the front of the cell as a 'sling' (white arrow) at $t = 3$ s (**a**) and 0.5 s (**b**). **c, d,** Three-dimensional reconstruction of the footprint of a rolling neutrophil. **c**, x–y plane and **d**, x–z plane. **e**, SEM of a rolling neutrophil showing tethers and slings. **f**, Unprocessed qDF image of a DiI-stained neutrophil rolling on P-selectin reveals wrapping of slings. **g**, SEM showing sling (red arrow) wrapped around a rolling neutrophil. **h**–**m**, Epifluorescence intravital microscopic images demonstrating sling formation by six different leukocytes. P-selectin (**a**–**g**), 20 molecules per μm^2. Shear stress, 10 dyn cm^{-2} (**a**, **c**–**g**), 8 dyn cm^{-2} (**b**). Arrowheads, tether anchor points; thin arrow, sling; thick arrow, rolling direction. Scale bars, 5 μm.

As each sling wraps around the rolling cell, each of the PSGL-1 patches functions as a discrete point of attachment with the P-selectin substrate (Fig. 3c, Supplementary Fig. 13 and Supplementary Movie 8). When a PSGL-1 patch on the sling becomes load-bearing, the bond force pulls the patch away from the substrate in the z-direction (Fig. 3d and Supplementary Fig. 14). Concomitantly, the apparent brightness of a PSGL-1 patch decreases by 50–75% (Fig. 3e and Supplementary Fig. 15) as the bonds get loaded because the intensity of the TIRF evanescent wave exponentially decreases with z-distance from the cover slip[4]. On tethers, PSGL-1 is concentrated in the tether anchorage point (Fig. 3b). When the PSGL-1 patch at the tether anchor point breaks, this results in catastrophic failure (step 1 in Fig. 3f), because there is no chance that the bonds can re-form. However, on slings, PSGL-1 is organized in patches and, unlike a tether, a sling peels in a step-wise manner (Fig. 3c, f and Supplementary Fig. 13). Figure 3g shows the impact of the failure of sling patches and tethers on the forward displacement of the cell shown in Fig. 3c. When a patch of PSGL-1 bonds fails, the sling is pulled by the rolling neutrophil until the next PSGL-1 patch becomes loaded. As a consequence, the forward displacement of the rolling cell following a patch failure is smaller (0.2 ± 0.1 (s.d.) μm; $n = 2$; Fig. 3g) than that of a tether failure (1 ± 0.3 (s.d) μm; $n = 4$; Fig. 3g), and the increase in displacement is short lived because a new bond patch is available to be loaded within less than 2 μm. The complete loss of a sling (between 2.2 and 2.4 s in Fig. 3g) is equivalent to the breaking of a tether, which results in a large forward displacement, because no new PSGL-1 patch is available to bear the load. A similar analysis of two more cells (Supplementary Fig. 16 and 17) show that the slings indeed act as breaks on rolling neutrophils. This was further corroborated by the *in vivo* observation (Supplementary Fig. 18) that rolling leukocytes slow down following sling formation in cremaster venules of mice. 'Step-wise peeling of slings' (white thin arrow in Supplementary Fig. 19) is distinct from 'pulling of tethers'[4,5] (white arrowhead in Supplementary Fig. 19). The two mechanisms probably work synergistically to facilitate shear-resistant rolling.

Under inflammatory conditions *in vivo*, mouse $\alpha_L\beta_2$-integrin LFA-1 binds ICAM-1 (ref. 15) and ICAM-2 (ref. 16) and is known to be responsible for neutrophil slow rolling[17] and arrest[18]. Unlike PSGL-1, LFA-1 is uniformly expressed on the surface of wild-type (Fig. 4a, b, Supplementary Figs 20 and 21, and Supplementary Movie 9) but not LFA-1-deficient neutrophils (Supplementary Fig. 22) and colocalizes with the footprint, tether anchorage points and slings (Fig. 4c). To show directly the differential distribution of PSGL-1 and LFA-1, we labelled the same cell suspension with both non-blocking-anti-PSGL-1 (red or green) and anti-LFA-1 (green or red) mAbs. Again, LFA-1 was seen in the footprint, tether anchorage points and slings, whereas PSGL-1 was restricted to microvilli tips, tether anchorage points and spots on slings (Fig. 4b and Supplementary Figs 23 and 24).

We have previously shown that neutrophils rolling on P-selectin acquire some LFA-1 in the extended conformation with intermediate affinity that can bind to ICAM-1 on the substrate to support a rolling interaction[17,19]. As slings are rich in LFA-1 and wrap around rolling

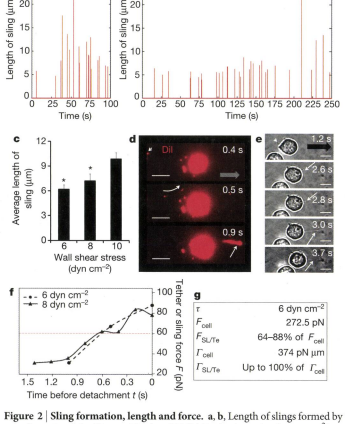


Figure 2 | Sling formation, length and force. **a**, **b**, Length of slings formed by a single neutrophil rolling on 20 (**a**) and 10 (**b**) P-selectin molecules per μm^2 in a microfluidic device. **c**, Average length of slings as a function of shear stress. Error bars, s.e.m.; $*P < 0.05$ relative to 10 dyn cm^{-2}; $n = 34$, 47, and 64 for 6, 8, and 10 dyn cm^{-2}, respectively. **d**, A tether (arrowhead at 0.4 s) swings to the front as a sling (white arrow at 0.9 s). qDF images processed. **e**, A tether (arrowhead at 1.2 s) swings to the front as a sling (white arrow at 3 s). DIC images. Shear stress 10 dyn cm^{-2}. P-selectin 20 molecules per μm^2. Scale bars, 5 μm. Rolling direction, thick arrow. **f**, Sling or tether force F before detachment from the substrate. Solid curve, 8 dyn cm^{-2}; dashed curve, 6 dyn cm^{-2}; solid symbols, qDF data. **g**, Fraction of the forward cell force and torque that can be balanced by 2 slings and 2 tethers. τ, shear stress; F_{cell}, total forward force; $F_{SL/Te}$, forward force balanced by 2 slings and 2 tethers; Γ_{cell}, total torque; $\Gamma_{SL/Te}$, torque balanced by 2 slings and 2 tethers. Data in **a**–**f** representative of five independent experiments.

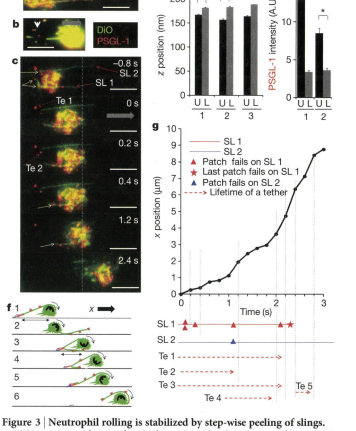


Figure 3 | Neutrophil rolling is stabilized by step-wise peeling of slings. **a**, PSGL-1 expressed in patches (red spots) on the sling (green; white arrow). **b**, PSGL-1 (red spots) concentrated in the tether anchorage points (white arrowhead). **c**, Two slings and anchorage points of two tethers denoted as SL 1 and SL 2 (red solid arrow) and Te 1 and Te 2 (red dashed arrow), respectively. PSGL-1 patches at the inflection point of SL 1 and SL 2 visible as red spots (solid white arrow) at ~0.8 s. As the cell rolls forward the SL 1 and SL 2 become load-bearing at 0 s, which is evident by the apparent loss of PSGL-1 staining (red) in these patches. The two load-bearing patches on SL 1 fail between 0 and 0.2 s and the next two downstream patches on SL 1 become load-bearing (red fluorescence in these patches reduced at 0.2 s). The last PSGL-1 patch on the SL 1 fails between 2.2 and 2.4 s. See Supplementary Fig. 13 for details. **d**, The z-distance of the sling membrane from the cover slip in three different PSGL-1 patches. **e**, PSGL-1 fluorescence intensity for two PSGL-1 patches. U, unloaded. L, loaded. $*P < 0.05$. Error bars, s.e.m. **f**, Schematic of step-wise peeling mechanism. Cell (green), PSGL-1 (red) and P-selectin (blue). **g**, Position of the cell shown in c plotted as a function of time. Closed circles, cell position; red and blue solid lines, SL 1 and SL 2, respectively; red and blue closed triangles, failure of individual patches on SL 1 and SL 2, respectively; red five-point star, failure of last patch on SL 1; red-dashed horizontal arrow, lifetime of a tether. Shear stress, 10 dyn cm^{-2}. P-selectin, 20 molecules per μm^2. Scale bars, 5 μm. Rolling direction, thick arrows. DqDF images (**a**–**c**) were processed. Data representative of six independent experiments.

neutrophils, we investigated whether slings can mediate LFA-1–ICAM-1 or LFA-1–ICAM-2 interactions in *trans* in rolling neutrophils. A recent study has shown that ICAM-1 is upregulated on a subset of mouse whole-blood neutrophils during ischaemia-reperfusion injury[20]. Flow cytometry (Supplementary Fig. 25a–c) showed that ICAM-1 is absent on blood and minimally expressed on bone marrow mouse neutrophils under the conditions used in this study. However, ICAM-2 was abundantly expressed on bone marrow neutrophils (Supplementary Fig. 25d–f). Bone marrow neutrophils from wild-type or LFA-1-deficient (*Itgal*$^{-/-}$) mice either untreated or blocked with function-blocking antibodies against LFA-1 (ref. 17) or ICAM-2 (ref. 21) were allowed to roll on P-selectin and the rolling velocities were compared (Fig. 4d). To our initial surprise, absence of LFA-1 or blocking LFA-1 or ICAM-2 resulted in a small but reproducible decrease in rolling velocity (Fig. 4d). As a positive control, bone marrow neutrophils from wild-type mice were also allowed to roll on glass substrate coated with both P-selectin and ICAM-1. This resulted in much larger decrease in rolling velocity (Supplementary Fig. 26), as shown previously[19]. As expected, treatment with function-blocking anti-LFA-1 mAb returned the rolling velocity to the level seen with LFA-1 blockade on a P-selectin substrate, but not to the higher level seen in neutrophils rolling on P-selectin without blockade of LFA-1 or ICAM-2 (Supplementary Fig. 26). DqDF images (Fig. 4e and Supplementary Fig. 27) show that although LFA-1 (red) is expressed both on cell surface and slings, ICAM-2 (green) is expressed mostly on the cell surface. These data indicate that the LFA-1 on slings interacts with ICAM-2 on the cell surface in *trans*. Green staining was absent in neutrophils stained with Alexa-Fluor-488-conjugated rat $IgG_{2a\kappa}$ (green) antibody (Supplementary Fig. 28). LFA-1 binding to ICAM-2 is expected to cause the sling to be more tightly wrapped around the rolling neutrophil, thus creating a less favourable lever arm (distance d in Fig. 4f). When the LFA-1–ICAM-2 interaction is absent, the lever arm becomes longer (distance d in Fig. 4g) and the rolling velocity decreases. The distance 'd' was indeed found to be significantly larger

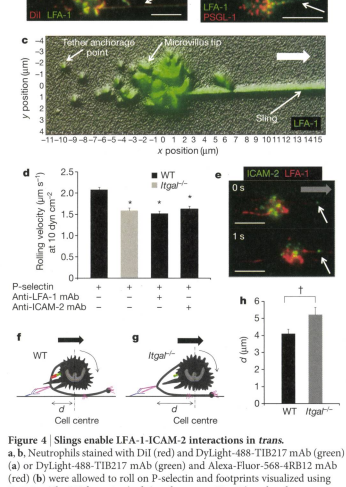


Figure 4 | Slings enable LFA-1-ICAM-2 interactions in *trans*. **a**, **b**, Neutrophils stained with DiI (red) and DyLight-488-TIB217 mAb (green) (**a**) or DyLight-488-TIB217 mAb (green) and Alexa-Fluor-568-4RB12 mAb (red) (**b**) were allowed to roll on P-selectin and footprints visualized using DqDF. **c**, The DiI footprint (red; Supplementary Fig. 20) analysed to generate a three-dimensional reconstruction and the LFA-1 image (green; Supplementary Fig. 20) overlaid to reveal distribution of LFA-1 (green) on sling, tether anchorage points, and microvilli (grey hills) versus cell surface (valleys). **d**, Mean rolling velocity of neutrophils on P-selectin. Wild type (WT) (black; $n = 159$), $Itgal^{-/-}$ (grey; $n = 173$), WT blocked with anti-LFA-1 mAb (black; $n = 142$), and WT blocked with anti-ICAM-2 mAb (black; $n = 98$). *$P < 0.01$ relative to WT without blocking. **e**, DqDF images of a rolling neutrophil show distribution of LFA-1 (red) and ICAM-2 (green). **f**, **g**, Schematics showing the mechanism of LFA-1-ICAM-2 interactions in *trans*. *d*, distance of the rearmost load-bearing patch at the inflection point of the sling from the cell centre; LFA-1 (red); ICAM-2 (green); PSGL-1 (violet); P-selectin (blue). **h**, Mean '*d*' plotted for WT ($n = 56$) and $Itgal^{-/-}$ ($n = 53$) neutrophils. †$P < 0.05$. Shear stress, 10 dyn cm^{-2}. P-selectin, 20 molecules per μm^2. Scale bars, 5 μm. White arrowheads, tether anchorage points; white arrows, slings; thick arrow, rolling direction. Error bars, s.e.m. Data representative of five (**a**–**c**), four (**d**) and three (**e**, **h**) experiments.

in LFA-1-deficient ($Itgal^{-/-}$) neutrophils compared to wild-type neutrophils (Fig. 4h and Supplementary Note 4).

In conclusion, the discovery of slings provides a mechanism by which neutrophils rolling at high shear stress pave their own way for enhanced rolling. The patchy distribution of PSGL-1 along each sling provides a unique adhesive substrate once the cell rolls over the sling. As each PSGL-1 patch fails, a new patch is already lined up that now becomes load-bearing. This step-wise peeling makes slings even more efficient than tethers in slowing down rolling neutrophils. Since slings accumulate over time, eventually, the rolling becomes very stable at high shear stresses[5]. Besides stabilizing rolling, slings are unique structures that also enable rolling neutrophils to present LFA-1 to its ligand ICAM-2 in *trans*. Taken together, catch bonds[22], long tethers[4], cell flattening[1,2,4] and now slings may explain why neutrophils can roll even at very high shear stress as observed in acute inflammation *in vivo*[1,3].

METHODS SUMMARY

The assembly of microfluidic devices and the coating of cover slips with recombinant murine P-selectin-Fc (20 molecules per μm^2) have been described previously[4,12]. In few experiments, cover slips coated with both recombinant murine P-selectin-Fc and ICAM-1-Fc were used in the microfluidic devices. Neutrophils were isolated from the bone marrow of mice using immune-magnetic negative selection and stained with Vybrant DiO or DiI (Molecular Probes) as described previously[12]. In some experiments, DiO- or DiI-stained neutrophils were additionally stained with Alexa-Fluor-568- or DyLight-488- or Alexa-Fluor-488-conjugated nonblocking-anti-mouse-PSGL-1 (4RB12) or function-blocking anti-mouse-LFA-1 (TIB217) or function-blocking anti-mouse ICAM-2 (3C4) mAb. Cells were suspended in RPMI-1640 (GIBCO-Invitrogen, Carlsbad, CA) + 10% mouse plasma at a concentration of 2.5×10^6 cells per ml and perfused through the microfluidic device at a wall shear stress of 6–10 dyn cm^{-2}. In few experiments, microfluidic devices were perfused with whole blood from anaesthetized *Lyz2-EGFP* mice[13] at a shear stress of 6–10 dyn cm^{-2}. The microfluidic device was placed on the stage of TIRF microscope (IX71; ×100 NA 1.45 Plan-Apochromatic oil immersion TIRFM objective; Olympus America) and images were recorded using qDI^2 or DqDF[12]. qDF images were processed and used to create three-dimensional reconstructions of the footprints as described previously[4,12]. In few experiments, neutrophils rolling in microfluidic devices were fixed with 2% paraformaldehyde and subjected to scanning electron microscopy. Mouse cremaster muscles were prepared for intravital microscopy and 25 μg of DyLight-488-conjugated TIB217 mAb was injected via the catheter connected to the right carotid artery to stain all leukocytes green. A glass cover slip (no. 1; 1 cm × 1 cm) was placed over the moistened cremaster muscle and epifluorescence observations were recorded using a ×100 NA 1.45 oil immersion objective on an Olympus BX61 upright microscope.

Full Methods and any associated references are available in the online version of the paper.

Received 14 December 2011; accepted 22 May 2012.
Published online 1 July 2012.

1. Firrell, J. C. & Lipowsky, H. H. Leukocyte margination and deformation in mesenteric venules of rat. *Am. J. Physiol. Heart Circ. Physiol.* **256,** H1667–H1674 (1989).
2. Sundd, P., Pospieszalska, M. K., Cheung, L. S., Konstantopoulos, K. & Ley, K. Biomechanics of leukocyte rolling. *Biorheology* **48,** 1–35 (2011).
3. Damiano, E. R., Westheider, J., Tozeren, A. & Ley, K. Variation in the velocity, deformation, and adhesion energy density of leukocytes rolling within venules. *Circ. Res.* **79,** 1122–1130 (1996).
4. Sundd, P. *et al.* Quantitative dynamic footprinting microscopy reveals mechanisms of neutrophil rolling. *Nature Methods* **7,** 821–824 (2010).
5. Ramachandran, V., Williams, M., Yago, T., Schmidtke, D. W. & McEver, R. P. Dynamic alterations of membrane tethers stabilize leukocyte rolling on P-selectin. *Proc. Natl Acad. Sci. USA* **101,** 13519–13524 (2004).
6. Schmidtke, D. W. & Diamond, S. L. Direct observation of membrane tethers formed during neutrophil attachment to platelets or P-selectin under physiological flow. *J. Cell Biol.* **149,** 719–730 (2000).
7. Shao, J.-Y. in *Leukocyte Adhesion* (ed. Ley, K.) Ch. 2, 25–45 (Academic, 2009).
8. Pospieszalska, M. K., Lasiecka, I. & Ley, K. Cell protrusions and tethers: a unified approach. *Biophys. J.* **100,** 1697–1707 (2011).
9. Hocde, S. A., Hyrien, O. & Waugh, R. E. Cell adhesion molecule distribution relative to neutrophil surface topography assessed by TIRFM. *Biophys. J.* **97,** 379–387 (2009).
10. Moore, K. L. *et al.* P-selectin glycoprotein ligand-1 mediates rolling of human neutrophils on P-selectin. *J. Cell Biol.* **128,** 661–671 (1995).
11. Waugh, R. E. in *Leukocyte Adhesion* (ed. Ley, K.) Ch. 1, 3–24 (Academic, 2009).
12. Sundd, P. *et al.* Live cell imaging of pauillin in rolling neutrophils by dual-color quantitative dynamic footprinting. *Microcirculation* **18,** 361–372 (2011).
13. Faust, N., Varas, F., Kelly, L. M., Heck, S. & Graf, T. Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. *Blood* **96,** 719–726 (2000).
14. Pendl, G. G. *et al.* Immature mouse dendritic cells enter inflamed tissue, a process that requires E- and P-selectin, but not P-selectin glycoprotein ligand 1. *Blood* **99,** 946–956 (2002).
15. Springer, T. A. Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. *Cell* **76,** 301–314 (1994).
16. Xu, H. *et al.* Characterization of murine intercellular adhesion molecule-2. *J. Immunol.* **156,** 4909–4914 (1996).
17. Zarbock, A., Lowell, C. A. & Ley, K. Spleen tyrosine kinase Syk is necessary for E-selectin-induced $\alpha_L\beta_2$ integrin-mediated rolling on intercellular adhesion molecule-1. *Immunity* **26,** 773–783 (2007).

18. Lawrence, M. B. & Springer, T. A. Leukocytes roll on a selectin at physiological flow rates: distinction from and prerequisite for adhesion through integrins. *Cell* **65,** 859–873 (1991).
19. Kuwano, Y., Spelten, O., Zhang, H., Ley, K. & Zarbock, A. Rolling on E- or P-selectin induces the extended but not high-affinity conformation of LFA-1 in neutrophils. *Blood* **116,** 617–624 (2010).
20. Woodfin, A. *et al.* The junctional adhesion molecule JAM-C regulates polarized transendothelial migration of neutrophils *in vivo*. *Nature Immunol.* **12,** 761–769 (2011).
21. Huang, M. T. *et al.* ICAM-2 mediates neutrophil transmigration *in vivo*: evidence for stimulus specificity and a role in PECAM-1-independent transmigration. *Blood* **107,** 4721–4727 (2006).
22. Marshall, B. T. *et al.* Direct observation of catch bonds involving cell-adhesion molecules. *Nature* **423,** 190–193 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank A. Zychlinsky for comments and reading the manuscript. This study was supported by the NCRP-Scientist Development Grant 11SDG7340005 from the American Heart Association (P.S.), WSA postdoctoral fellowship 10POST4160142-01 from American Heart Association (E.K.K.) and NIH EB 02185 (K.L.).

Author contributions P.S. performed all the experiments and image analysis. E.G. and A.G. designed the microfluidic device. M.K.P. calculated the fraction of bond force and torque shared by slings and tethers. E.K.K. was involved in culturing of Th1 CD4 T cells. Y.K. and S.F performed the scanning electron microscopy. P.S. and K.L. wrote the manuscript. K.L. supervised the project. All authors discussed the results and commented on the manuscript.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.L. (klaus@liai.org).

LETTER

doi:10.1038/nature11262

A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response

Laurens Kruidenier[1]*, Chun-wa Chung[2]*, Zhongjun Cheng[3]*, John Liddle[1]*, Ka Ming Che[4,5]*, Gerard Joberty[6], Marcus Bantscheff[6], Chas Bountra[4], Angela Bridges[2], Hawa Diallo[1], Dirk Eberhard[6], Sue Hutchinson[2], Emma Jones[2], Roy Katso[2], Melanie Leveridge[2], Palwinder K. Mander[1], Julie Mosley[2], Cesar Ramirez-Molina[1], Paul Rowland[2], Christopher J. Schofield[4], Robert J. Sheppard[1], Julia E. Smith[1], Catherine Swales[5], Robert Tanner[2], Pamela Thomas[2], Anthony Tumber[4], Gerard Drewes[6], Udo Oppermann[4,5], Dinshaw J. Patel[3], Kevin Lee[1]† & David M. Wilson[1]

The jumonji (JMJ) family of histone demethylases are Fe^{2+}- and α-ketoglutarate-dependent oxygenases that are essential components of regulatory transcriptional chromatin complexes[1–4]. These enzymes demethylate lysine residues in histones in a methylation-state and sequence-specific context[5]. Considerable effort has been devoted to gaining a mechanistic understanding of the roles of histone lysine demethylases in eukaryotic transcription, genome integrity and epigenetic inheritance[2,4,6], as well as in development, physiology and disease[3,7]. However, because of the absence of any selective inhibitors, the relevance of the demethylase activity of JMJ enzymes in regulating cellular responses remains poorly understood. Here we present a structure-guided small-molecule and chemoproteomics approach to elucidating the functional role of the H3K27me3-specific demethylase subfamily (KDM6 subfamily members JMJD3 and UTX)[8]. The liganded structures of human and mouse JMJD3 provide novel insight into the specificity determinants for cofactor, substrate and inhibitor recognition by the KDM6 subfamily of demethylases. We exploited these structural features to generate the first small-molecule catalytic site inhibitor that is selective for the H3K27me3-specific JMJ subfamily. We demonstrate that this inhibitor binds in a novel manner and reduces lipopolysaccharide-induced proinflammatory cytokine production by human primary macrophages, a process that depends on both JMJD3 and UTX. Our results resolve the ambiguity associated with the catalytic function of H3K27-specific JMJs in regulating disease-relevant inflammatory responses and provide encouragement for designing small-molecule inhibitors to allow selective pharmacological intervention across the JMJ family.

Initially, we applied a structure–function approach to elucidate the underlying recognition principles that account for the exquisite specificity of JMJD3 for H3K27me3 (trimethylated lysine 27 on histone 3) over other histone peptides containing the common ARKS amino acid motif[9]. We solved high-resolution crystal structures of the approximately 500 carboxy-terminal amino acids of mouse and human JMJD3 in the presence of cofactor and metal (Fig. 1a, Supplementary Fig. 1a and Supplementary Tables 1 and 2). The orthologue structures superimpose well on each other. The crystal structure of the histone H3(20–34)K27me3 peptide bound to an engineered mouse JMJD3 construct in the presence of the cofactor analogue *N*-oxalylglycine (NOG) and metal (Ni^{2+}) at 2.5 Å resolution is shown in Fig. 1b. This truncated JMJD3 protein encompasses residues 1,157 to 1,641 and includes a cleaved insertion segment that replaces a disordered loop, to facilitate crystal packing and peptide binding (Supplementary Fig. 2a, b).

The JMJD3 enzyme is composed of a JmjC catalytic domain and a C-terminal segment within which is embedded a Zn^{2+}-coordinated GATA-like (GATAL) domain of novel topology that is flanked by α-helical segments that form a four-helix bundle (Fig. 1b and Supplementary Fig. 1b, c). The two domains pack against each other with a large buried surface area (4,115 $Å^2$), explaining the requirement for the C-terminal domain for optimal stability and catalytic competence of the truncated JMJD3 protein (Supplementary Fig. 3).

The H3K27me3 peptide binds in a channel within the catalytic domain and amino acid residues A24 to G34 make several hydrogen-bonding and van der Waals interactions with the enzyme as is evident in Fig. 1c (see also Supplementary Fig. 4). The K27me3 group inserts deep into the catalytic pocket, bringing it close to the NOG and Ni^{2+} that lie at its base (Fig. 1d). The hydrogen-bonding interactions that anchor the H3K27 residue are shown in Supplementary Fig. 5a. The methylene side chain of K27me3 forms hydrophobic contacts with M1373. Elsewhere, the catalytic pocket is lined with polar residues: asparagine, glutamine, serine and tyrosine. The structure shows that the histone peptide resides both amino- and C-terminal to K27, especially P30 and R26, make several interactions (such as with R1246, P1388 and E1244 of JMJD3) that are important for specificity and recognition.

Both H3K9 and H3K27 have a common ARKS sequence context. It has been shown that H3K9me2 and H3K27me3 peptides bind with the same directionality to KDM7A, a dual-specificity H3K9/H3K27 demethylase from *Caenorhabditis elegans*[9] (Supplementary Fig. 6a, b). The present study shows that the H3K27me3 peptide binds to the H3K27-specific demethylase JMJD3 in the opposite orientation to KDM7A. In part, this difference is associated with P30, which points inwards in the complex with JMJD3 but outwards in the complex with KDM7A (Supplementary Fig. 6b, c). In addition, a comparison of intermolecular contacts in our H3K27me3–JMJD3 structure with the corresponding structures of H3K9me2–PHF8 (ref. 10) and H3K9me3–JMJD2A[11,12] establishes that residues outside the ARKS segment also contribute to substrate specificity (Supplementary Fig. 7).

To understand the substrate recognition and to determine the impact on lysine demethylase activity, we mutated key residues of the H3K27me3 peptide–enzyme complex, as determined from the crystal structure (Supplementary Figs 8 and 9). In the bound state of the H3 peptide, R26 forms hydrogen bonds directly with E1244 and D1333 (Fig. 1c). The importance of these hydrogen bonds is shown by the lack of substrate turnover of the R26A mutant peptide, in which the arginine at position 26 has been replaced with alanine (Fig. 1e and Supplementary Fig. 8). Mutation of the P30 residue of H3, which is directed towards P1388 of JMJD3 (Fig. 1c), to glycine, serine or phenylalanine resulted in a loss of activity (Fig. 1e). The P30A mutant

[1]Epinova DPU, Immuno-Inflammation Therapy Area, GlaxoSmithKline R&D, Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK. [2]Platform Technology and Science, GlaxoSmithKline R&D, Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK. [3]Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [4]Structural Genomics Consortium, University of Oxford, Old Road Campus, Roosevelt Drive, Headington OX3 7DQ, UK. [5]Botnar Research Centre, NIHR Biomedical Research Unit, University of Oxford OX3 7LD, UK. [6]Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany. †Present address: Pfizer, Biotherapeutics R&D, 200 Cambridgepark Drive, Cambridge, Massachusetts 02140, USA.
*These authors contributed equally to this work.

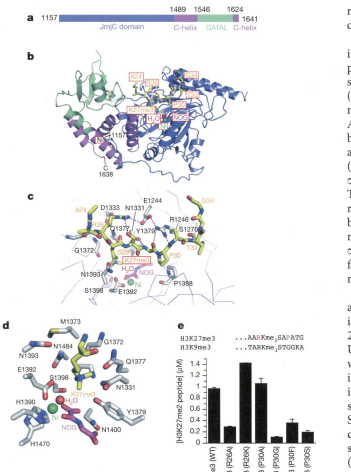


Figure 1 | Structure of the H3K27me3-derived peptide bound to mouse JMJD3. a, Domain topology and numbering system of JMJD3 construct used for crystallization of the enzyme in the free state. **b**, Overview of the structure of H3(20–34)K27me3–mouse JMJD3(1157–1641) complex, with the histone peptide (K27me3, yellow) in a stick representation and the engineered enzyme (Supplementary Fig. 2) colour coded by domain as in **a**: GATAL (green), JmjC (blue) and helical (pink) domains. The heteroatoms in the stick representation are coloured according to atom type: oxygen (red) and nitrogen (blue). The N and C termini of the enzyme are indicated. **c**, View of the intermolecular interactions between the bound histone peptide (yellow) and the enzyme residues (grey), with hydrogen bonds shown as dashed pink lines. **d**, View of the catalytic pocket of the enzyme (grey), the base of which is lined by NOG (pink) and Ni^{2+} (green). The histone peptide is shown in yellow. Details of the hydrogen bonding in the catalytic pocket are outlined in Supplementary Fig. 5a. **e**, Effect of mutations of the native H3K27me3 substrate peptide (highlighted in red) at R26 and P30 on turnover. Mass spectrometric quantification of the H3K27me2 product was performed after a 6-min reaction with concentrations of peptide spanning 0.6 to 250 μM (Supplementary Fig. 8). Data are shown at approximately the K_m (Michaelis constant) of the standard, wild-type (WT), H3K27me3 peptide (22 μM) for clarity. Data are presented as the mean ± s.d.

retained activity, highlighting the requirement for an appropriate hydrophobic interaction with P1388 (Fig. 1e). It has been observed that serine/threonine kinases such as MSK1 can phosphorylate H3S28 (ref. 13). Interestingly, JMJD3 did not demethylate the H3K27me3 peptide when it was phosphorylated at S28 (H3K27me3S28ph) (Supplementary Fig. 9). Our findings also indicate that the JMJD3 residues E1244, R1246, D1333 and P1388 are likely to be important for substrate recognition, as single mutations of these residues to alanine caused a loss of lysine demethylation activity. Structural and sequence comparisons of JMJD3, UTX and UTY suggest that these recognition determinants are preserved within the KDM6 subfamily of demethylases[14] (Supplementary Figs 10–12).

We exploited these structural insights, in particular the protein interactions made by α-ketoglutarate and the P30 residue of the H3 peptide, to optimize a series of weakly active hits discovered from a screening of the GlaxoSmithKline corporate compound collection (~2 million compounds[15]). The resultant lead, GSK-J1, had a half-maximum inhibitory concentration (IC_{50}) of 60 nM in the JMJD3 AlphaScreen assay (Fig. 2a). A 1.8 Å co-crystal structure of GSK-J1 bound to human JMJD3 revealed the critical interactions within the active site (Fig. 2b, Supplementary Table 2 and Supplementary Fig. 13 (omit electron-density map)). The propanoic acid of GSK-J1 mimics α-ketoglutarate binding by maintaining interactions with K1381, T1387 and N1480 (Fig. 2b and Supplementary Fig. 14a). The aromatic ring of the tetrahydrobenzazepine of GSK-J1 sits in a narrow cleft between R1246 and P1388 (Fig. 2b and Supplementary Fig. 14b), mimicking P30 of the histone peptide. GSK-J1 is competitive with α-ketoglutarate but non-competitive with the peptide substrate. This finding suggests that GSK-J1's mechanism does not follow a simple random model for bisubstrate inhibition[16] (Supplementary Fig. 15).

The pyridyl-pyrimidine biaryl of GSK-J1 makes a bidentate interaction with the catalytic metal (Fig. 2b) and induces a shift in the Co^{2+} ion (used in this crystallization experiment to mimic the Fe^{2+} ion) of 2.34 Å away from the conserved HHE triad (Supplementary Fig. 16). Uniquely, this movement results in the metal cation swapping position with a previously apical water molecule. Subsequently, H1470 makes an indirect water-bridged interaction with the metal, whereas the direct interactions of the metal with H1390 and E1392 are preserved through small changes in their side-chain conformations (Fig. 2c and Supplementary Fig. 14c). The discovery of this subtle but distinctive dynamic metal shift in JMJD3 provides a further opportunity to derive selectivity by designing compounds that take advantage of this elasticity (Supplementary Fig. 17 and Supplementary Information (Materials)).

The overlay of the peptide and inhibitor complexes (Fig. 2c and Supplementary Fig. 18) highlights the shift in the divalent cation, as well as the overlap between bound GSK-J1 and the NOG binding site, thereby accounting for cofactor displacement in the inhibitor complex and the observed mechanism of action.

GSK-J1 is selective for the H3K27 demethylases JMJD3 and UTX and is inactive against a panel of demethylases of the JMJ family, as measured by a combination of thermal shift, mass spectrometry[17] and antibody-based assays[18] (Fig. 3a, b, Supplementary Fig. 19 and Supplementary Table 3). GSK-J1 also did not significantly inhibit 100 protein kinases at a concentration of 30 μM in a competition-binding assay (Supplementary Table 4) and had negligible off-target activity against a panel of 60 unrelated proteins, including other chromatin-modifying enzymes such as histone deacetylases (Supplementary Table 5).

The bidentate interaction between GSK-J1 and the catalytic metal is critical for enzyme inhibition. The pyridine regio-isomer GSK-J2 (Fig. 2a) cannot form such an interaction and hence showed considerably weaker JMJD3 inhibition ($IC_{50} > 100$ μM, Supplementary Table 3). Because GSK-J2 has analogous physicochemical properties to GSK-J1 but lacks H3K27 demethylase activity, it provides an ideal inactive control molecule for elucidating the functional role of JMJD3 inhibition.

Given the unusual binding mode of GSK-J1, we investigated its selectivity and its ability to engage full-length JMJD3 and UTX in an endogenous cellular environment by using a chemoproteomics approach. This approach combines a competition-binding assay in a cell extract, using an immobilized analogue of the inhibitor, with western blotting and mass spectrometry analysis[19]. The co-crystal structure of GSK-J1 with JMJD3 indicated that substitution at the *para* position to the pyridine nitrogen would enable access to solvent and would hence be amenable to immobilization, yielding GSK-J3 (Fig. 3c). GSK-J3 retained good activity against JMJD3 and was attached to sepharose beads to generate a KDM6 probe matrix. This

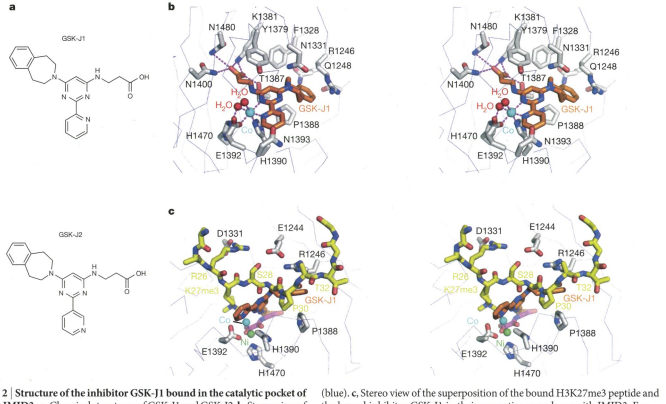


Figure 2 | Structure of the inhibitor GSK-J1 bound in the catalytic pocket of human JMJD3. a, Chemical structures of GSK-J1 and GSK-J2. **b**, Stereo view of the intermolecular interactions between the bound inhibitor GSK-J1 (orange) and the human JMJD3 enzyme residues (grey) that line the catalytic pocket, with hydrogen bonds shown as dashed pink lines. The heteroatoms in the stick representation are coloured according to atom type: oxygen (red) and nitrogen (blue). **c**, Stereo view of the superposition of the bound H3K27me3 peptide and the bound inhibitor GSK-J1 in their respective complexes with JMJD3. For clarity, the histone peptide (yellow), NOG (pink), Ni^{2+} (green) and certain JMJD3 side chains (grey) are shown only from the peptide complex, and GSK-J1 (orange) and Co^{2+} (blue) are shown from the inhibitor complex.

matrix precipitated Flag-tagged full-length JMJD3 and UTX from transiently transfected HEK-293 cells (Fig. 3d). This interaction with JMJD3 and UTX was specific, because the addition of free GSK-J1 to the lysate completely inhibited the binding of JMJD3 and UTX to the probe matrix (Fig. 3d). To confirm the ability of GSK-J1 to bind to endogenous JMJD3 in a relevant physiological context, we next used phorbol myristate acetate (PMA)-stimulated HL-60 monocytic cells to induce the expression of JMJD3, which was then specifically captured by the GSK-J3 probe matrix (Fig. 3e). Furthermore, in quantitative mass spectrometric experiments, JMJD3 was the only protein that was specifically captured by the probe matrix, as indicated by the inhibition of its binding in the presence of free GSK-J1 (Fig. 3f), thus confirming the remarkable selectivity of GSK-J1 in a cellular context.

Although critical for *in vitro* binding (Fig. 2b), the highly polar carboxylate group of GSK-J1 restricts cellular permeability. Therefore, we used a pro-drug strategy, masking the polarity of the acid groups of GSK-J1 and GSK-J2 with ethyl esters and yielding GSK-J4 and GSK-J5 (Fig. 4a). These cell-penetrating esters (with *in vitro* potency in the mass spectrometry assay, $IC_{50} > 50\ \mu M$) are rapidly hydrolysed by macrophage esterases, thereby generating pharmacologically relevant intracellular concentrations of GSK-J1 and GSK-J2 (Table 1). The ethyl ester pro-drug GSK-J4 was confirmed to have cellular activity in Flag–JMJD3-transfected HeLa cells, in which GSK-J4 prevented the JMJD3-induced loss of nuclear H3K27me3 immunostaining (Fig. 4b and Supplementary Fig. 20). Administration of GSK-J4 increased total nuclear H3K27me3 levels in untransfected cells (Supplementary Fig. 20c).

JMJD3 is involved in several physiological functions, including the inflammatory response[20–23]. In macrophages, JMJD3 expression is rapidly induced by proinflammatory stimuli through a nuclear factor-κB (NF-κB)-dependent mechanism, and JMJD3 is recruited to the transcription start sites (TSSs) of over 70% of lipopolysaccharide (LPS)-induced genes, where it participates directly in the transcriptional response[24,25]. Crucially, it remains unclear whether this activation of transcription is achieved through the demethylation of H3K27me3 at target gene promoters[24]. Therefore, we next examined the efficacy of GSK-J4 and GSK-J5 at inhibiting the LPS-induced response of human primary macrophages derived from healthy volunteers. Administration of GSK-J4 significantly reduced the expression of 16 of 34 LPS-driven cytokines as assessed by PCR array (Fig. 4c and Supplementary Table 6), including tumour-necrosis factor-α (TNF-α).

TNF-α is important in a variety of inflammatory disorders, prompting us to make a more detailed characterization of GSK-J4's control of this cytokine. TNF-α protein production was inhibited in a dose-dependent manner (IC_{50} of GSK-J4 for TNF-α blockade, 9 μM), whereas the inactive isomer, GSK-J5, had no effect (Fig. 4d). No interference from cellular toxicity was observed under these conditions (Supplementary Fig. 21a); upstream NF-κB signalling seemed to be intact (Fig. 4e); and neither GSK-J4 nor GSK-J5 directly affected the levels of LPS-induced JMJD3 or of LPS-independent UTX and EZH2 (Supplementary Fig. 21b). Kinetic analysis suggested that the early phases of TNF-α induction were less affected by our inhibitor than were the later phases (Supplementary Fig. 21c, d). Interestingly, short interfering RNA (siRNA)-mediated knockdown of either JMJD3 or UTX alone in human primary macrophages had no effect on TNF-α production. Under these conditions, GSK-J4 was still able to inhibit TNF-α production, suggesting redundancy within the KDM6 subfamily in the regulation of this cytokine. To mimic the pan-KDM6 activity of GSK-J4, we performed dual depletion of JMJD3 and UTX in these cells, and this depletion eliminated TNF-α production (Fig. 4f and Supplementary Fig. 22). In addition, chromatin immunoprecipitation (ChIP) studies confirmed that GSK-J4, but not GSK-J5, prevented the LPS-induced loss of H3K27me3 associated with the *TNFA* TSS and blocked the recruitment of RNA polymerase II to this locus (Fig. 4g). Neither GSK-J4 nor GSK-J5 had an effect on H3K27 methylation

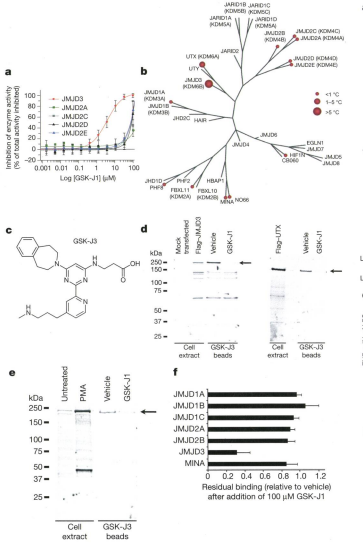


Figure 3 | GSK-J1 is selective for H3K27 demethylases of the KDM6 subfamily and specifically binds to endogenous JMJD3. **a**, Evaluation of the selectivity of GSK-J1 in JMJ mass spectrometric assays. Data are presented as the mean ± s.d. **b**, Phylogenetic tree of human JMJ enzymes illustrating the selectivity of GSK-J1 for demethylases of the KDM6 subfamily over other KDM subfamilies of methyl-lysine demethylases, as determined by melting point (T_{m}) shift screening. Temperature differences (°C) are shown in red circles. Only UTX and JMJD3 show significant stabilization (>1 °C), indicating ligand binding. **c**–**f**, Chemoproteomics studies show target engagement in a cellular environment. **c**, GSK-J3, an amine analogue of GSK-J1 used for bead immobilization. **d**, Immobilized GSK-J3 was able to affinity-capture Flag-tagged JMJD3 and Flag-tagged UTX from transiently transfected HEK-293 cells. Western blotting analysis with an anti-Flag antibody shows the capture of JMJD3 (left, arrow) and UTX (right, arrow), both of which were competitively inhibited by the addition of 100 μM GSK-J1. **e**, PMA treatment of HL-60 cells (80 nM for 16 h) induced the expression of JMJD3, as shown by western blotting. Immobilized GSK-J3 captured the endogenous JMJD3 protein (arrow), and this capture was competitively inhibited by 100 μM GSK-J1. **f**, GSK-J1 is selective for JMJD3, as shown by mass-spectrometry-based quantification of JMJC proteins (captured with GSK-J3 beads) from PMA-stimulated HL-60 cells in the presence of 100 μM GSK-J1. Data are presented as the mean from two experiments, and error bars represent the 95% confidence intervals.

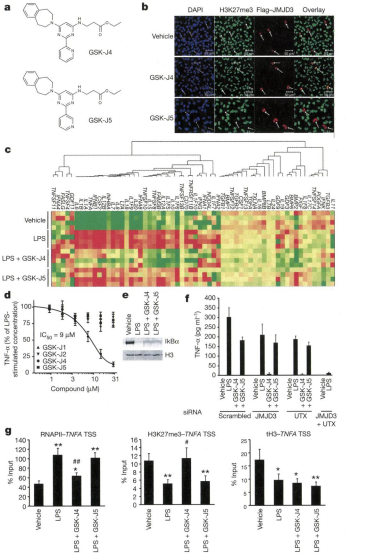


Figure 4 | GSK-J1 inhibits TNF-α production by human primary macrophages in an H3K27-dependent manner. **a**, Chemical structures of the ethyl ester pro-drugs GSK-J4 and GSK-J5. **b**, Administration of 25 μM GSK-J4, but not GSK-J5, preserved nuclear H3K27me3 staining (green) in Flag–JMJD3-transfected (red, arrows) HeLa cells. Scale bars, 50 μm. **c**, Heat map representation of cytokine expression by human primary macrophages activated with LPS (for 2 h) in the presence of 30 μM GSK-J4 or GSK-J5 ($n = 4$ donors; green, low expression; red, high expression). **d**, TNF-α production by human primary macrophages activated with LPS in the presence of the indicated concentrations of GSK-J4 or GSK-J5 for 6 h (data are presented as the mean ± s.e.m. from $n = 5$ donors). **e**, Western blot showing the degradation of inhibitor of NF-κB α (IκBα) in whole cell lysates of LPS-stimulated human primary macrophages (30 min) in the presence of 30 μM GSK-J4 or GSK-J5. **f**, TNF-α production by LPS-stimulated human primary macrophages (1 h) transfected with scrambled, or JMJD3-directed or UTX-directed siRNA, in the presence of 30 μM GSK-J4 or GSK-J5 (data are presented as the mean ± s.d. from five transfection replicates from one representative experiment). **g**, ChIP analysis of the association of RNA polymerase II (RNAPII), H3K27me3 and total H3 (tH3) with the *TNFA* TSS in LPS-stimulated human primary macrophages (1 h) in the presence of 30 μM GSK-J4 or GSK-J5 (data are presented as the mean ± s.e.m. from $n = 8$ donors). *P* values were calculated using a two-tailed Student's *t*-test. *, $P < 0.05$ compared with vehicle; **, $P < 0.01$ compared with vehicle; #, $P < 0.05$ compared with LPS; ##, $P < 0.01$ compared with LPS.

upstream of the *TNFA* TSS (Supplementary Fig. 23). Finally, GSK-J4, but not GSK-J5, blocked the production of TNF-α by macrophages derived from patients with rheumatoid arthritis (Supplementary Fig. 24).

Our findings provide a strategy for the rational design and characterization of catalytic site inhibitors of JMJ enzymes. Specifically, we exploited active site plasticity and targeted key substrate-binding features to identify the first series of selective H3K27 inhibitors. A chemoproteomics approach was used to assess the engagement of endogenous targets and to confirm the remarkable selectivity of the inhibitors within a relevant cellular context. By using these inhibitors, we provide clear insight into the biological relevance of the demethylase

Table 1 | Intracellular delivery of GSK-J1 and GSK-J2

Compound	[GSK-J1] in lysate (μM)	[GSK-J2] in lysate (μM)
Acid (GSK-J1 or GSK-J2)	1.6 ± 1.6	7.8 ± 1.6
Ethyl ester (GSK-J4 or GSK-J5)	11.8 ± 0.6	17.4 ± 1.5

GSK-J1 and GSK-J2 were detected in the lysates of human primary macrophages 1 h after the administration of 30 μM GSK-J1 or GSK-J2 or 30 μM of the pro-drugs GSK-J4 or GSK-J5 (mean ± s.d. from $n = 3$ donors).

function of the KDM6 JMJ enzymes. Furthermore, we show that H3K27 demethylation and specifically JMJD3 and UTX catalytic activity are critical determinants of proinflammatory gene activation in human primary macrophages. In reporting the first selective KDM6 inhibitor, we show the relevance and tractability of demethylase inhibition, as well as indicate that the KDM6 class of proteins can be targeted by epigenetic focused drug discovery, which may have broad therapeutic application.

METHODS SUMMARY

The cloning, expression, protein purification and crystallization of mouse JMJD3(1132–1641), engineered mouse JMJD3(1157–1641)–H3, human JMJD3 and their complexes are described in the Supplementary Information. GSK-J1–GSK-J5 were synthesized as pure compounds and fully characterized as outlined in the Supplementary Information. The fluorescent thermal shift assays, AlphaScreen, and RapidFire and matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometric assays to measure the inhibition of histone demethylase activity and JMJ selectivity, as well as the chemoproteomics experiments to establish compound–target engagement, were carried out as described in the Supplementary Information or as previously reported[19]. The procedures for the studies in human primary macrophages, including intracellular compound quantification, quantitative PCR with reverse transcription, enzyme-linked immunosorbent assays, western blotting, siRNA-mediated knockdown of JMJD3 and UTX, and ChIP are outlined in the Supplementary Information, together with a description of the antibodies and the materials.

Received 30 September 2011; accepted 28 May 2012.
Published online 29 July 2012.

1. Mosammaparast, N. & Shi, Y. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu. Rev. Biochem.* **79,** 155–179 (2010).
2. Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. *Nature Rev. Mol. Cell Biol.* **8,** 307–318 (2007).
3. Shi, Y. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nature Rev. Genet.* **8,** 829–833 (2007).
4. Trojer, P. & Reinberg, D. Histone lysine demethylases and their impact on epigenetics. *Cell* **125,** 213–217 (2006).
5. Loenarz, C. & Schofield, C. J. Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem. Sci.* **36,** 7–18 (2011).
6. Shi, Y. & Whetstine, J. R. Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell* **25,** 1–14 (2007).
7. Pedersen, M. T. & Helin, K. Histone demethylases in development and disease. *Trends Cell Biol.* **20,** 662–671 (2010).
8. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
9. Yang, Y. *et al.* Structural insights into a dual-specificity histone demethylase ceKDM7A from *Caenorhabditis elegans*. *Cell Res.* **20,** 886–898 (2010).
10. Horton, J. R. *et al.* Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nature Struct. Mol. Biol.* **17,** 38–43 (2010).
11. Couture, J. F. *et al.* Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nature Struct. Mol. Biol.* **14,** 689–695 (2007).
12. Ng, S. S. *et al.* Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448,** 87–91 (2007).
13. Gehani, S. S. *et al.* Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. *Mol. Cell* **39,** 886–900 (2010).
14. Sengoku, T. & Yokoyama, S. Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev.* **25,** 2266–2277 (2011).
15. Gamo, F. J. *et al.* Thousands of chemical starting points for antimalarial lead identification. *Nature* **465,** 305–310 (2010).
16. Yu M., *et al.* Bisubstrate inhibition: theory and application to *N*-acetyltransferases. *Biochemistry* **45,** 14788–14794 (2006).
17. Mackeen, M. M. *et al.* Small-molecule-based inhibition of histone demethylation in cells assessed by quantitative mass spectrometry. *J. Proteome Res.* **9,** 4082–4092 (2010).
18. Sakurai, M. *et al.* A miniaturized screen for inhibitors of Jumonji histone demethylases. *Mol. Biosyst.* **6,** 357–364 (2010).
19. Bantscheff, M. *et al.* Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nature Biotechnol.* **29,** 255–265 (2011).
20. Lan, F. *et al.* A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* **449,** 689–694 (2007).
21. Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318,** 447–450 (2007).
22. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449,** 731–734 (2007).
23. Miller, S. A., Mohn, S. E. & Weinmann, A. S. Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Mol. Cell* **40,** 594–605 (2010).
24. De Santa, F. *et al.* Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. *EMBO J.* **28,** 3341–3352 (2009).
25. De Santa, F. *et al.* The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* **130,** 1083–1094 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the following colleagues for their help and advice in preparing the manuscript: A. Argyrou, S. Atkinson, M. Barker, M. Campbell, L. Gordon, R. Gregory, R. Grimley, P. Humphreys, P. Jeffrey, A. Kucik, B. Leavens, C. Lewis, V. Ludbrook, R. Prinjha, S. Ratcliffe, C. Sharp, D. Thomas, D. Tough and M. Woodrow. Help was also provided by N. Burgess-Brown, S. Ng, J. Dunford, A. Kawamura, N. Rose and M. Daniel. D.J.P. was supported by funds from the Starr Foundation, the Abby Rockefeller Mauze Trust and the Maloris Foundation. The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canada Foundation for Innovation, Genome Canada, GlaxoSmithKline, Lilly Canada, the Novartis Research Foundation, Pfizer, Takeda, the Ontario Ministry of Economic Development and Innovation, and the Wellcome Trust. This study was supported by the National Institute for Health and Research (NIHR) Biomedical Research Unit, Oxford. Thanks also to M. Boesche and the Cellzome biochemistry and mass spectrometry teams for assistance with the chemoproteomics experiments. Finally, we acknowledge additional contributions from members of Platform Technology Sciences and the Epinova DPU and the JMJD3 programme team for supporting this research.

Author Contributions L.K., C.-w.C., J.L., Z.C., G.D., U.O., D.J.P., K.C., K.L. and D.M.W. defined the research, interpreted the data and wrote the manuscript. All authors performed the experiments, analysed the data and/or contributed to the scientific discussion.

Author Information Crystal structures of JMJD3 have been deposited in the Protein Data Bank under accession numbers 4EZ4 (mouse JMJD3–NOG–Ni^{2+}), 4EYU (mouse JMJD3–NOG–Ni^{2+}), 4EZH (mouse JMJD3–H3K27me3–Ni^{2+}), 2XUE (human JMJD3–α-ketoglutarate–Fe^{2+}) and 4ASK (human JMJD3–GSK-J1–Co^{2+}). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.M.W. (david.m.wilson@gsk.com).

LETTER

doi:10.1038/nature11272

The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming

Abed AlFatah Mansour[1]*, Ohad Gafni[1]*, Leehee Weinberger[1], Asaf Zviran[1], Muneef Ayyash[2], Yoach Rais[1], Vladislav Krupalnik[1], Mirie Zerbib[1], Daniela Amann-Zalcenstein[3,4], Itay Maza[1], Shay Geula[1], Sergey Viukov[1], Liad Holtzman[1], Ariel Pribluda[1], Eli Canaani[5], Shirley Horn-Saban[4], Ido Amit[3], Noa Novershtern[1]* & Jacob H. Hanna[1]

Induced pluripotent stem cells (iPSCs) can be derived from somatic cells by ectopic expression of different transcription factors, classically Oct4 (also known as Pou5f1), Sox2, Klf4 and Myc (abbreviated as OSKM)[1]. This process is accompanied by genome-wide epigenetic changes[2–5], but how these chromatin modifications are biochemically determined requires further investigation. Here we show in mice and humans that the histone H3 methylated Lys 27 (H3K27) demethylase Utx[6–9] (also known as Kdm6a) regulates the efficient induction, rather than maintenance, of pluripotency. Murine embryonic stem cells lacking Utx can execute lineage commitment and contribute to adult chimaeric animals; however, somatic cells lacking Utx fail to robustly reprogram back to the ground state of pluripotency. Utx directly partners with OSK reprogramming factors and uses its histone demethylase catalytic activity to facilitate iPSC formation. Genomic analysis indicates that Utx depletion results in aberrant dynamics of H3K27me3 repressive chromatin demethylation in somatic cells undergoing reprogramming. The latter directly hampers the derepression of potent pluripotency promoting gene modules (including Sall1, Sall4 and Utf1), which can cooperatively substitute for exogenous OSK supplementation in iPSC formation. Remarkably, Utx safeguards the timely execution of H3K27me3 demethylation observed in embryonic day 10.5–11 primordial germ cells (PGCs)[10], and Utx-deficient PGCs show cell-autonomous aberrant epigenetic reprogramming dynamics during their embryonic maturation *in vivo*. Subsequently, this disrupts PGC development by embryonic day 12.5, and leads to diminished germline transmission in mouse chimaeras generated from Utx-knockout pluripotent cells. Thus, we identify Utx as a novel mediator with distinct functions during the re-establishment of pluripotency and germ cell development. Furthermore, our findings highlight the principle that molecular regulators mediating loss of repressive chromatin during *in vivo* germ cell reprogramming can be co-opted during *in vitro* reprogramming towards ground state pluripotency.

We established primed epiblast stem cells (EpiSCs)[11,12] carrying a Nanog–GFP reporter and doxycycline-inducible OSKM alleles, and applied piggyBac (PB)-based gene-trap screening[13] to identify inhibitors of EpiSC reversion to naive pluripotency (Supplementary Fig. 1a–c). Several clonal populations analysed carried a disruptive PB integration in the second intron of the *Utx* gene and failed to generate Nanog–GFP$^+$ naive pluripotent cells adequately (Fig. 1a and Supplementary Fig. 1d). Functional rescue by overexpression of Utx induced rapid Nanog–GFP reactivation (Fig. 1a).

Utx is an X-chromosome-encoded gene that, together with *Uty* and *Jmjd3* (also known as Kdm6b), belongs to a small family of JmjC-domain containing enzymes that mediate demethylation of H3K27 tri- and dimethyl (me3/2) repressive chromatin marks[6–9]. We knocked-out *Utx* in V6.5 mouse embryonic stem (ES) cell line and examined their pluripotent characteristics (Supplementary Fig. 2). V6.5 $Utx^{\Delta/Y}$ ES cells were pluripotent, as evidenced by normal staining for pluripotency markers, global gene expression, and growth rate and lineage commitment marker upregulation upon differentiation (Fig. 1b and Supplementary Figs 3, 4). Whereas Utx-deficient embryos predominantly die around embryonic day (E)10.5 (ref. 14), and at E8.5–E9.5 following tetraploid complementation assay (Supplementary Fig. 5), viable high-contribution adult chimaeras were readily obtained from $Utx^{\Delta/Y}$ ES cells (Fig. 1c and Supplementary Fig. 4). These results demonstrate that *Utx* is dispensable for pluripotent state maintenance, and that functional commitment into multiple somatic lineages can be achieved in the absence of this gene.

We next tested whether Utx may be required for the re-establishment, rather than maintenance, of pluripotency. Thus, we knocked out *Utx* in two different 'secondary reprogramming' pluripotent stem cell lines (NGFP1 iPSCs and V19 ES cells) carrying doxycycline-inducible OSKM transgenes[15]. Genetically matched NGFP1 or V19 $Utx^{\Delta/Y}$ and $Utx^{+/Y}$ cells were injected into host blastocysts, and differentiated transgenic mouse embryonic somatic fibroblasts (MEFs) and pre-B cells were isolated. Induction of pluripotency in these cells by doxycycline demonstrated a marked block in retrieving iPSCs from $Utx^{\Delta/Y}$ cells in naive and primed reprogramming conditions (Fig. 1d and Supplementary Fig. 6). Long-term follow-up[15,16] of reprogramming of monoclonal populations of donor somatic $Utx^{\Delta/Y}$ B cells showed a severely delayed detection of Nanog–GFP reactivation following doxycycline induction (Supplementary Fig. 7). Notably, reprogrammed Nanog–GFP$^+$ cells that could be derived at low efficiency and delayed kinetics retained aberrant global H3K27me3 and H3K4me3 epigenetic profiles (Supplementary Fig. 8), consistent with aberrant reprogramming. Human fibroblasts demonstrated reduced iPSC formation upon UTX knockdown (Supplementary Fig. 9). Furthermore, the lack of Utx inhibited the epigenetic reversion of primed EpiSCs and reprogramming via cell fusion (Supplementary Fig. 10). These results demonstrate that the absence of Utx demethylase constitutes a major roadblock for re-establishing pluripotency *in vitro*.

Utx is ubiquitously expressed in somatic tissues (Supplementary Fig. 5b), and its ectopic overexpression in addition to OSKM does not lead to increased efficiency in iPSC formation (Supplementary Fig. 11a, b). Acute tamoxifen-induced ablation of *Utx* in NGFP1-CreER; $Utx^{flox/Y}$ transgenic reprogrammable cells at different partially reprogrammed stages[17] resulted in diminished progression towards pluripotency (Supplementary Fig. 11c). However, once the cells had reactivated the endogenous Nanog–GFP reporter (characterized to constitute the final molecular barrier to pluripotency[18]), deleting *Utx* with tamoxifen had no influence on obtaining iPSCs *in vitro* (Supplementary Fig. 11c). These results indicate that Utx expression

[1]The Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]The Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel. [3]The Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel. [4]The Department of Biological Services—Genomics Unit, Weizmann Institute of Science, Rehovot 76100, Israel. [5]The Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel.
*These authors contributed equally to this work.

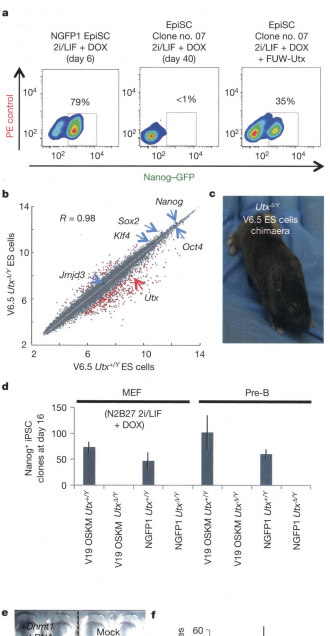


Figure 1 | Utx promotes the re-establishment of pluripotency in somatic cells. a, Fluorescence-activated cell sorting (FACS) analysis for Nanog–GFP reporter conducted on the indicated EpiSC-derived cell samples and mutant clones grown in 2i/LIF (2i/LIF is media containing LIF and two small-molecule inhibitors: PD0325901 (ERK1/2 inhibitor) and CHIR99021 (GSK-3β inhibitor)) plus doxycycline (DOX) conditions. PE, phycoerythrin. **b**, Scatter plot of transcriptional levels measured in $Utx^{+/Y}$ (*x*-axis, $n = 1$) and $Utx^{d/Y}$ (*y*-axis, $n = 1$) ES cells. Transcription levels are shown in $\log_2$ scale. Rank correlation between the samples is shown. **c**, Adult chimaeria as indicated by the agouti coat colour. **d**, Reprogramming of 5,000 secondary OSKM-transgenic MEF or pre-B donor cells after 16 days on doxycycline in naive growth conditions, and quantifying $Nanog^+$ iPSC colonies ($n = 3$). N2B27, serum-free media (see Methods). **e**, $Utx^{d/Y}$ somatic cells on doxycycline for 7 days were transduced with different rescue overexpression vectors and analysed 7 days later after passage for iPSC formation via staining for alkaline phosphatase. **f**, As in **e**, except rescue was conducted by *Eed* siRNA transfection ($n = 2$). Error bars indicate s.d.

is not a rate-limiting factor for iPSC formation, but is rather essential and continuously required during *in vitro* reprogramming. Specific knockdown of Jmjd3 did not significantly reduce iPSC formation (Supplementary Figs 9, 12). Furthermore, Utx-deficient somatic cell reprogramming could not be reconstituted by overexpression of JMJD3, UTY or other tested pluripotency promoting factors, suggesting a unique requirement for Utx in this process (Fig. 1e). Importantly, $Utx^{d/Y}$ iPSC formation rescue was observed by overexpressing wild-type UTX, but not the UTX^{H1146A} allele, which carries a mutation in the JmjC catalytic domain, abrogating its H3K27 demethylase activity (Fig. 1e)[7]. We next tested whether inducing global loss of H3K27me1/2/3 could facilitate the reprogramming of $Utx^{d/Y}$ somatic cells. Thus, we inhibited Eed expression, a component of the polycomb group protein repressive complex (PRC2) required for the establishment and maintenance of H3K27me marks, and its inhibition leads to genome-wide loss of H3K27me1/2/3 (refs 16, 19). Short interfering RNA (siRNA) for Eed resulted in the rapid derivation of iPSCs from $Utx^{d/Y}$ somatic cells undergoing reprogramming following doxycycline induction (Fig. 1f). This indicates that Utx predominantly contributes to reprogramming towards naive pluripotency through its H3K27 demethylation activity, and that authentic reprogramming towards the ground state of pluripotency correlates with global loss of H3K27me3 deposition. However, our results do not exclude synergistic contributions from Jmjd3, or demethylation-independent functions of Utx in reprogramming.

We next conducted gene expression profiling of somatic cells during reprogramming and compared them to patterns observed in ES cells. The effect of *Utx* deletion on the induction, rather than on the maintenance, of the pluripotent state was evident at the transcriptional level. The number of genes affected by *Utx* knockout was 6.6 times higher in MEF samples after doxycycline induction, compared to ES cells (Supplementary Fig. 13a and Supplementary Table 1). Hierarchical clustering and principle component analysis clustered $Utx^{d/Y}$ doxycycline-induced MEFs together with untreated $Utx^{d/Y}$ and $Utx^{+/Y}$ MEFs (Fig. 2a and Supplementary Fig. 13b), thus indicating that $Utx^{d/Y}$ MEFs fail to initiate molecular changes typically detected after 8 days of OSKM induction, and remain highly similar to the original donor somatic cells (Pearson correlation = 0.97). Importantly, the genes that are activated (>2-fold) during reprogramming in $Utx^{+/Y}$ MEFs are enriched (false discovery rate (FDR) <1%) for targets of Klf4, Oct4 and Sox2 factors, as was measured in ES cells or during iPSC formation after OSKM induction in MEFs[20] (Fig. 2b). In contrast, known targets of the latter reprogramming factors were preferentially enriched among the downregulated gene subgroup in $Utx^{d/Y}$ doxycycline-induced MEFs (>2-fold) (Fig. 2b). In both cases, the affected genes were highly enriched for H3K27me3 chromatin marks that were aberrantly retained in $Utx^{d/Y}$ MEFs following doxycycline induction (Supplementary Table 2).

We next mapped H3K27me3 and H3K4me3 chromatin modifications by genome-wide chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) analysis in MEFs before and after doxycycline induction in genetically matched $Utx^{+/Y}$ and $Utx^{d/Y}$ samples and in established ES cells (Supplementary Table 3). The perturbed active H3K4me3 and repressive H3K27me3 chromatin marks correlated with global active and repressed transcriptional levels, respectively[2] (Supplementary Fig. 13c). The effect of *Utx* deletion during iPSC reprogramming was also apparent at the global chromatin level. Twenty-seven per cent of the 14,619 mapped genes had aberrant H3K27me3 and/or H3K4me3 modifications in $Utx^{d/Y}$ cells (Fig. 2c). A total of 957 genes retained or aberrantly gained H3K27me3 repressive marks (Fig. 2c) during reprogramming in

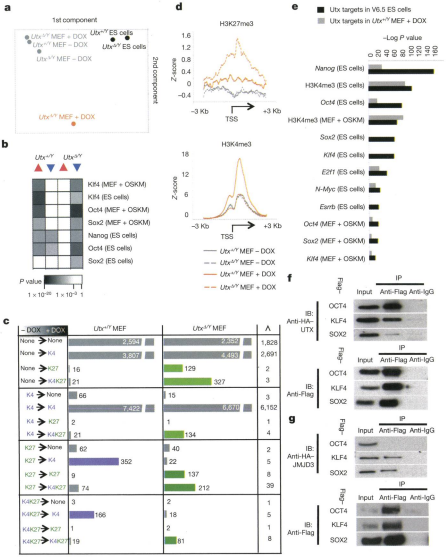


Figure 2 | Genomic dissection of reprogramming in Utx-knockout cells. **a**, Principle component analysis of transcriptional profiles of the indicated samples. Colours indicate distinct clusters: ES cells (black); MEF and MEF-like (grey); and doxycycline (DOX)-induced partially reprogrammed pre-iPSCs (orange). **b**, Selected functional categories enriched in induced (red) and repressed (blue) genes (>2-fold), following 8 days of doxycycline in $Utx^{+/Y}$ and $Utx^{\Delta/Y}$ MEFs. Grey intensity indicates P value. **c**, Number of genes with H3K4me3 or H3K27me3 before and after doxycycline induction in $Utx^{+/Y}$ and $Utx^{\Delta/Y}$ MEFs. Light green indicates number of genes that gained H3K27me3 in $Utx^{\Delta/Y}$ only; dark green indicates number of gene that retained H3K27me3 in $Utx^{\Delta/Y}$ only; purple indicates number of gene that gained/maintained H3K4me3 in $Utx^{+/Y}$ only. Right column shows overlapping gene number in each category. **d**, Binding profile around the transcription start site (TSS) of H3K4me3 and H3K27me3. Profiles were calculated over 805 genes that are bound by at least 3 of the following factors: Oct4, Sox2, Klf4, Myc and Nanog. Z-score of the mean profile is plotted (y-axis) against a sliding window around the TSS (x-axis). **e**, Top categories enriched among genes bound by Utx in ES cells (black) or in MEFs + DOX (grey), following genome-wide ChIP-seq analysis conducted either on ES cells or MEFs after OSKM transduction (as indicated in brackets). −Log P value is presented for each category. **f**, **g**, Co-imuunoprecipoitation (IP) of Flag-tagged OCT4, KLF4 or SOX2 with HA-tagged UTX or JMJD3, followed by immunoblot analysis (IB).

$Utx^{\Delta/Y}$ and not in $Utx^{+/Y}$ MEFs. This suggests that Utx also safeguards against aberrant acquisition of repressive H3K27me3 marks after OSKM induction[2]. More than 500 genes became potentially active by losing H3K27me3 and gaining or maintaining H3K4me3 active chromatin marks in $Utx^{+/Y}$, but not in $Utx^{\Delta/Y}$ doxycycline-induced MEFs (Fig. 2c). Notably, *Utx* deletion resulted in global increase in H3K27me3 in doxycycline-induced MEFs but not in ES cells (Supplementary Fig. 14a). This trend is consistent with our functional data indicating that Utx predominantly regulates pluripotency re-establishment rather than maintenance. H3K4me3 was decreased in doxycycline-induced MEFs but not in ES cells (Supplementary Fig. 14a), possibly resulting from an indirect effect of perturbation in H3K27me3-repressed genes and/or from the fact that Utx partners with the Mll2/3–H3K4me3-transferase complex[6,7]. Consistent with the gene-expression enrichment analysis in Fig. 2b, known OSKM and Nanog directly bound target genes in ES cells and MEFs undergoing reprogramming[20] were found to be methylated for H3K27me3 and demethylated for H3K4me3 in $Utx^{\Delta/Y}$ doxycycline-induced MEFs (Fig. 2d). Overall, the above analysis provides evidence for aberrant transcriptional and epigenetic reprogramming dynamics upon OSKM induction in *Utx*-knockout somatic cells, and suggests that genes regulated by the pluripotency reprogramming machinery cannot be adequately reactivated in the absence of Utx.

ChIP-seq analysis for Utx binding in ES cells identified 2,416 bound genes that are enriched for H3K4me3 active chromatin marked genes as well as for targets of Oct4, Sox2 and Klf4 (Fig. 2e). Motif search analysis in Utx-binding locations identified predicted binding motifs for Klf4, Sox2, Oct4 and Nanog as the four most highly enriched (Supplementary Fig. 14b). ChIP-seq analysis for Utx binding in doxycycline-induced MEFs showed that Utx specifically bound 1,845 target genes that are enriched for H3K4me3 active chromatin marked genes as well as for targets of Nanog and Oct4 (Fig. 2e). Utx targets in MEFs following doxycycline induction, aberrantly accumulated H3K27me3 marks in these loci in $Utx^{\Delta/Y}$ cells (Supplementary Fig. 14c). The above findings prompted us to examine whether Utx physically associates with the OSK reprogramming proteins. We established that Flag-tagged OCT4, KLF4 and SOX2 specifically co-immunoprecipitated with haemagglutin (HA)-tagged UTX (Fig. 2f). JMJD3 was not able to interact directly with OCT4, partially explaining the distinct influence for different H3K27 demethylases on iPSC formation (Fig. 2g).

We then aimed to delineate unbiasly some of the key genes that are deregulated early during the reprogramming process by the loss of Utx. We identified 1,430 genes that were transcriptionally upregulated (>2-fold) in $Utx^{+/Y}$ and not in $Utx^{\Delta/Y}$ MEFs undergoing reprogramming. Of these, 98 genes had aberrantly upregulated H3K27me3

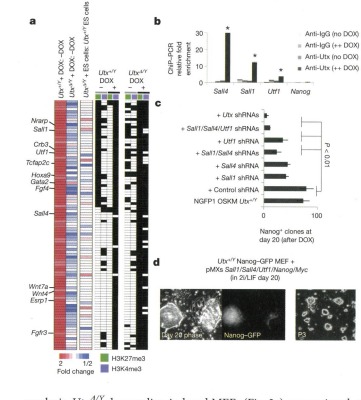


Figure 3 | Downstream effectors of Utx that promote iPSC reprogramming. **a**, Unbiased identification of 98 deregulated genes that were induced (>2-fold) in $Utx^{+/Y}$ but not in $Utx^{\Delta/Y}$. Left, expression change before and after 8 days of doxycycline (DOX) induction in $Utx^{+/Y}$ and $Utx^{\Delta/Y}$ MEFs, and in ES cells (comparing $Utx^{\Delta/Y}$ to $Utx^{+/Y}$). Right, chromatin marks before and after doxycycline induction. Purple indicates H3K4me3 modifications; green indicates H3K27me3 modifications; black boxes, methylated; white boxes, unmethylated. **b**, Anti-Utx and control anti-IgG ChIP–PCR relative fold enrichment of *Sall4*, *Sall1*, *Utf1* and *Nanog* loci in $Utx^{+/Y}$ and $Utx^{+/Y}$ MEFs before and after doxycycline treatment (normalized to $Utx^{+/Y}$ without doxycycline) ($n = 3$). **c**, Reprogramming following individual and combined knockdown with *Sall4*, *Sall1* and *Utf1* were introduced in $Utx^{+/Y}$ Nanog–GFP secondary MEFs ($n = 2$). *P value < 0.01 in comparison to wild-type samples. Error bars indicate s.d. **d**, Overexpression of *Sall4*, *Sall1*, *Utf1*, *Nanog* and *Myc* reproducibly gives rise to Nanog–GFP$^+$ iPSCs in $Utx^{+/Y}$ cells without exogenous OSK. P3, passage 3.

marks in $Utx^{\Delta/Y}$ doxycycline-induced MEFs (Fig. 3a), suggesting that their deregulation is likely to be caused by H3K27me3 retention. The signature of these deregulated genes is enriched (FDR <1%) for stem-cell maintenance factors (for example, *Fgf4*, *Sall4*, *Sall1* and *Utf1*), for validated direct targets of Klf4 and Oct4 (for example, *Sall1*, *Crb3* and *Nef1*), and for H3K27me3-marked genes (Supplementary Table 4). Of these, we focused on three transcription factors, *Sal1l*, *Sall4* and *Utf1*, previously shown to undergo reactivation after OSKM induction in wild-type cells[2]. These genes were directly bound by Utx in ES cells and in doxycycline-induced MEFs, and aberrantly retained H3K27me3 in $Utx^{\Delta/Y}$ cells, as was confirmed by ChIP-seq and ChIP–polymerase chain reaction (PCR) analyses (Fig. 3b and Supplementary Figs 14d, 15–16a). This suggested that Utx can act during early reprogramming by directly regulating the reactivation of a variety of pluripotency promoting factors including Sall1, Sall4 and Utf1. Indeed, inhibiting the endogenous expression of *Sall1*, *Sall4* and *Utf1* in $Utx^{+/Y}$ MEFs through specific knockdown during reprogramming, recapitulates the block in iPSC formation observed in $Utx^{\Delta/Y}$ cells (Fig. 3c). Overexpression of Sall1, Sall4 and Utf1 in $Utx^{\Delta/Y}$ MEFs promoted the activation of early reprogramming marker SSEA1 (Supplementary Fig. 16b), but did not result in reactivation of Nanog–GFP, suggesting that other Utx-regulated pluripotency factors need to be reactivated to complete the reprogramming progression. This is consistent with the role of Utx as global regulator that targets multiple genes (Fig. 2e). Remarkably, consistent with a profound pluripotency promoting potential for Sall1, Sall4 and Utf1, their combined overexpression with Nanog and Myc in wild-type $Utx^{+/Y}$ MEFs replaces exogenous OSK expression, and generated chimaera-formation-competent iPSC clones (Fig. 3d). Overall, these data

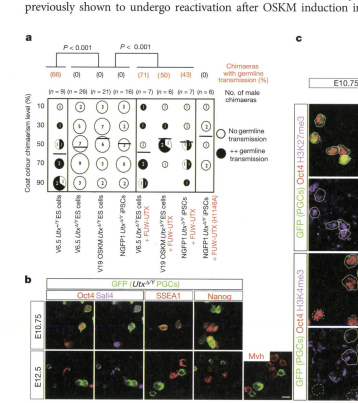


Figure 4 | Utx regulates epigenetic reprogramming and development of PGCs. **a**, Quantification of coat colour chimaerism in male mice derived from the indicated cell lines. Each pie chart indicates a group with the indicated chimaerism levels. Black areas indicate succesful germline transmission, white areas indicate germline transmission failure. *P* values are indicated. **b**, **c**, Confocal images represent gonad sections from E10.75 or E12.5 chimaeric embryos injected with $Utx^{\Delta/Y}$ BVSC ES cells and immunostained for the indicated proteins. **b**, Immunostaining for GFP (labelling BVSC $Utx^{\Delta/Y}$ injected cells) and the indicated PGC markers revealed abnormal *in vivo* differentiation of Utx-deficient PGCs at E12.5 but not at E10.75. **c**, Immunofluorescence staining for epigenetic markers on the genital ridge of E10.75 embryos demonstrates that high levels of H3K27me3 are retained in Utx-deficient PGCs (unbroken circles) compared with adjacent wild-type PGCs (dashed circles). At E12.5 Utx-deficient PGCs have diminished levels of all tested epigenetic marks. Scale bars, 10 μm.

indicate that Utx can associate with OKS by a protein–protein interaction and promotes the reactivation of potent pluripotency promoting modules that cooperatively facilitate iPSC formation.

We next investigated whether Utx has a role in epigenetic reprogramming during early development. Whereas ES cells could be readily derived from $Utx^{A/Y}$ blastocysts (Supplementary Fig. 17), we noticed that chimaeric animals generated from three different $Utx^{A/Y}$ iPSC and ES cell clones reproducibly failed to demonstrate germline transmission after extensive breeding (Fig. 4a). Importantly, multiple adult chimaeras generated from three different corrected transgenic lines with catalytically active UTX, but not the mutant UTX^{H1146A} allele, robustly achieved germline transmission (Fig. 4a). This indicates that the defect in germ-lineage development can be unambiguously assigned to the absence of catalytically active Utx.

PGCs increase global levels of H3K27me3 by E8.5–E9.5, and then undergo a second wave of epigenetic reprogramming at E10.5–11 that entails global loss of the H3K27me3 mark[10]. Thus we tested whether Utx deficiency results in perturbed PGC development and reprogramming *in vivo* during this critical developmental window. To reproducibly study PGC development beyond E9.5–E10.5, we analysed chimaeric animals after microinjection of either V6.5 $Utx^{A/Y}$ ES cells labelled with LacZ, or the BVSC ES cell line[21] knocked out for *Utx* ($Utx^{A/Y}$; BVSC), which carries Blimp1–mVenus (Blimp1 also known as Prdm1) and Stella–CFP (Stella also known as Dppa3) early PGC reporters. $Utx^{A/Y}$ PGCs were found to express the PGC markers Oct4, Nanog, Sall4 and SSEA1 normally at E10.75 (Fig. 4b). However, only *Utx*-null PGCs did not undergo H3K27me3 demethylation during this developmental window (Fig. 4c and Supplementary Fig. 18). The remaining chromatin marks tested were not different at this stage, suggesting that aberrantly accumulated H3K27me3 at this critical developmental stage may be an early identifiable trigger for subsequent PGC developmental defects. Indeed, despite the adequate localization of Blimp1- and Stella-positive $Utx^{A/Y}$ PGCs in the gonads at E12.5, they fail to retain the expression of several PGC pluripotency markers (Nanog, Sall4, Oct4 and SSEA1) or show reactivation of the late PGC marker Mvh (Fig. 4b and Supplementary Figs 18, 19). In addition, $Utx^{A/Y}$ PGCs show aberrant chromatin dynamics at E12.5, as evident from the low chromatin levels of H3K9me3, H3K4me2/3 and H3K27me3, and reduced nuclear size in comparison to adjacent wild-type PGCs (Fig. 4c and Supplementary Fig. 18c). Finally, we find that $Utx^{A/Y}$ PGCs isolated from E8.5 chimaeras are deficient in their ability to revert to ground state pluripotency and form ES-like embryonic germ cells (Supplementary Figs 20, 21). Taken together, these results provide direct evidence for cell-autonomous aberrant germ cell development *in vivo* and perturbed epigenetic reprogramming in the absence of Utx.

In summary, we identify Utx as a critical regulator acting at molecular switches during reprogramming to ground state pluripotency to safeguard an efficient, timely and authentic execution of H3K27 demethylation that is critical for this process. Future experimental work is needed to contrast the developmental potency phenotypes in mice and pluripotent cells deficient for different combinations of components regulating the deposition of H3K27 and H3K4 methylation (for example, Utx, Jmjd3, Uty and Mll2/3 enzymes).

METHODS SUMMARY

The following knockout stem cell lines for Utx were generated by targeted disruption of the endogenous *Utx* locus via homologous recombination: (1) V6.5 ES cells; (2) NGFP1 iPSCs; (3) V19-OSKM ES cells; and (4) BVSC ES cells[21]. Full details of stem cell lines, cell culture, plasmids, siRNAs and antibodies used, as well as descriptions of methods for epigenetic reprogramming, immunofluorescence, immunoblotting, immunoprecipitation, embryo micromanipulation, PGC differentiation, bio-informatics and statistical analyses are provided in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 6 February; accepted 31 May 2012.
Published online 8 July 2012.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Koche, R. P. *et al.* Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8,** 96–105 (2011).
3. Hemberger, M., Dean, W. & Reik, W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nature Rev. Mol. Cell Biol.* **10,** 526–537 (2009).
4. Singhal, N. *et al.* Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* **141,** 943–955 (2010).
5. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).
6. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in *HOX* gene regulation and development. *Nature* **449,** 731–734 (2007).
7. Hong, S. *et al.* Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc. Natl Acad. Sci. USA* **104,** 18439–18444 (2007).
8. Sen, G. L., Webster, D. E., Barragan, D. I., Chang, H. Y. & Khavari, P. A. Control of differentiation in a self-renewing mammalian tissue by the histone demethylase JMJD3. *Genes Dev.* **22,** 1865–1870 (2008).
9. Lan, F. *et al.* A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* **449,** 689–694 (2007).
10. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452,** 877–881 (2008).
11. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448,** 196–199 (2007).
12. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448,** 191–195 (2007).
13. Guo, G., Huang, Y., Humphreys, P., Wang, X. & Smith, A. A PiggyBac-based recessive screening method to identify pluripotency regulators. *PLoS ONE* **6,** e18189 (2011).
14. Lee, S., Lee, J. W. & Lee, S. K. UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. *Dev. Cell* **22,** 25–37 (2011).
15. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462,** 595–601 (2009).
16. Montgomery, N. D. *et al.* The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr. Biol.* **15,** 942–947 (2005).
17. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).
18. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).
19. Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biol.* **10,** 1291–1300 (2008).
20. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136,** 364–377 (2009).
21. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146,** 519–532 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.H.H. is supported by a generous gift from Ilana and Pascal Mantoux, The Sir Charles Clore Research Prize, an ERC starting investigator grant (StG-2011-281906), the Israel Science Foundation Regular and Bikura research grants, the ICRF Foundation, Fritz Thyssen Stiftung, the Alon Foundation scholar award, a grant from E. A. and R. Drake, and the Leona M. and Harry B. Helmsley Charitable Trust. N.N. and A.P. are supported by Weizmann Dean fellowship awards. We thank A. Smith, H. Niwa and M. Saitou for reagents. We thank the Weizmann Institute management for providing critical financial and infrastructural support.

Author Contributions O.G., A.A.M., N.N. and J.H.H. concieved the idea for this project, designed and conducted experiments and wrote the manuscript. L.W. conducted PGC differentiation. A.Z. conducted computational simulations. N.N. conducted bioinformatics analysis. O.G., I.A., D.A.-Z., S.H.-S. and E.C. conducted genomic and chromatin immunoprecipiation experiments. A.A.M., A.P., L.H. and I.M. conducted teratoma and embryo analysis. O.G. and S.V. generated knockout cells. J.H.H., O.G., V.K. and Y.R. conducted reprogramming experiments. M.Z. conducted microinjections. S.G. conducted pre-implantation analysis. M.A. provided critical technical advice.

Author Information Microarray data are available at the National Center for Biotechnology Information Gene Expression Omnibus database under the series accession number GSE37822. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.H.H. (jacob.hanna@weizmann.ac.il) or N.N. (noa.novershtern@weizmann.ac.il).

LETTER

doi:10.1038/nature11267

The dynamic disulphide relay of quiescin sulphydryl oxidase

Assaf Alon[1], Iris Grossman[1], Yair Gat[1], Vamsi K. Kodali[2], Frank DiMaio[3], Tevie Mehlman[4], Gilad Haran[5], David Baker[3], Colin Thorpe[2] & Deborah Fass[1]

Protein stability, assembly, localization and regulation often depend on the formation of disulphide crosslinks between cysteine side chains. Enzymes known as sulphydryl oxidases catalyse *de novo* disulphide formation and initiate intra- and intermolecular dithiol/disulphide relays to deliver the disulphides to substrate proteins[1,2]. Quiescin sulphydryl oxidase (QSOX) is a unique, multi-domain disulphide catalyst that is localized primarily to the Golgi apparatus and secreted fluids[3] and has attracted attention owing to its overproduction in tumours[4,5]. In addition to its physiological importance, QSOX is a mechanistically intriguing enzyme, encompassing functions typically carried out by a series of proteins in other disulphide-formation pathways. How disulphides are relayed through the multiple redox-active sites of QSOX and whether there is a functional benefit to concatenating these sites on a single polypeptide are open questions. Here we present the first crystal structure of an intact QSOX enzyme, derived from a trypanosome parasite. Notably, sequential sites in the disulphide relay were found more than 40 Å apart in this structure, too far for direct disulphide transfer. To resolve this puzzle, we trapped and crystallized an intermediate in the disulphide hand-off, which showed a 165° domain rotation relative to the original structure, bringing the two active sites within disulphide-bonding distance. The comparable structure of a mammalian QSOX enzyme, also presented here, shows further biochemical features that facilitate disulphide transfer in metazoan orthologues. Finally, we quantified the contribution of concatenation to QSOX activity, providing general lessons for the understanding of multi-domain enzymes and the design of new catalytic relays.

The introduction of disulphide bonds into folding proteins is typically accomplished by enzyme pairs[1,2]; one enzyme generates disulphides *de novo*, often with the aid of a bound cofactor, and the second enzyme acquires these disulphides by dithiol/disulphide exchange and passes them on to substrate proteins. Examples of such partnerships include DsbB and DsbA in the *Escherichia coli* periplasm and, in eukaryotes, Ero1 and protein disulphide isomerase (PDI) in the endoplasmic reticulum and Erv1 and Mia40 in the mitochondrial intermembrane space. A partnership between vitamin K epoxide reductase (VKOR) and membrane-anchored PDI-family proteins has also been identified in the endoplasmic reticulum[6]. The QSOX enzymes[3] are the only known example of conserved concatenation of disulphide-generating and disulphide-transferring modules within a single polypeptide (Fig. 1a), although a VKOR-like protein is fused to its partner in a few species[7]. Domain fusion may allow QSOX to function under dilute conditions in extracellular environments[8–10]; other disulphide-generating enzymes function in sequestered intracellular compartments with high local concentrations of disulphide-transferring enzymes.

Formation and transfer of disulphide bonds in QSOX are mediated by redox-active cysteine pairs in the canonical pattern Cys-X-X-Cys (Fig. 1a). One such dicysteine motif is in a thioredoxin-fold domain (Trx1) related to the redox-active domains of PDI. Another is in a flavin adenine dinucleotide (FAD)-binding domain (Erv) related to the mitochondrial enzyme Erv1. The proposed QSOX mechanism[11,12] involves the generation of disulphides catalytically at the FAD-proximal dicysteine motif, intramolecular disulphide transfer to the Trx1 domain and finally intermolecular dithiol/disulphide exchange with the substrate (Fig. 1a). How the Trx1 redox-active residues interact alternately with the Erv domain and with substrate proteins remains unclear.

The *Trypanosoma brucei* parasite, the cause of African trypanosomiases, encodes a QSOX enzyme (TbQSOX)[13]. We determined the X-ray crystal structure of TbQSOX to 2.3 Å resolution (Supplementary Table 1). In the crystallized configuration, the Trx1 redox-active site faces outwards (Fig. 1b), where it would be readily accessible to nucleophilic attack by substrate cysteine thiolates. However, in this position, the Trx1 active site is 42 Å away from the Erv active site and would thus be unable, after disengaging from the substrate, to accomplish the next step in the proposed relay mechanism (Fig. 1a).

To test whether the distance between redox-active sites in the TbQSOX crystal structure is representative of their average separation in solution, we used fluorescence resonance energy transfer (FRET) spectroscopy. TbQSOX Cys72 was mutated to leave one cysteine (Cys69) free within the Trx1 active site for labelling with the donor fluorophore Pacific Blue C_5-maleimide. The bound FAD, although not fluorescent, served as an energy acceptor at the Erv active site, and the decrease in donor fluorescence was taken to reflect FRET efficiency. The use of an intrinsic chromophore as one member of the FRET pair minimized potential perturbations due to labelling. An average distance of 55 Å was measured between the label at position 69 and FAD (Supplementary Table 2), supporting a requirement for major conformational changes during the QSOX disulphide relay. Donor labels introduced at a variety of other positions across the enzyme surface provided a set of ensemble-averaged distances characterizing the resting state of the enzyme (Fig. 1c, Supplementary Figs 1–3 and Supplementary Tables II and IIIa–c).

To obtain structural insight into the intramolecular disulphide-transfer step (see Supplementary Discussion), we eliminated the second cysteine of both TbQSOX redox-active Cys-X-X-Cys motifs. The resulting mutant is able to form, but not resolve, the interdomain disulphide essential for catalysis. We purified (Supplementary Fig. 4) and crystallized the closed, interdomain disulphide-bonded version of TbQSOX ($TbQSOX_C$) and solved its structure to 3.3 Å resolution. When compared with the wild-type enzyme, a ~165° rotation of the Trx1 domain and associated helix brings the two redox-active sites within covalent-bonding distance in $TbQSOX_C$ (Fig. 1d and Supplementary Movie 1). Consistent with this closed configuration being a transient state during enzyme turnover, the surface complementarity between the $TbQSOX_C$ redox-active domains is greater than

[1]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, USA. [3]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [4]Department of Biological Research Support, Weizmann Institute of Science, Rehovot 76100, Israel. [5]Department of Chemical Physics, Weizmann Institute of Science, Rehovot 76100, Israel.

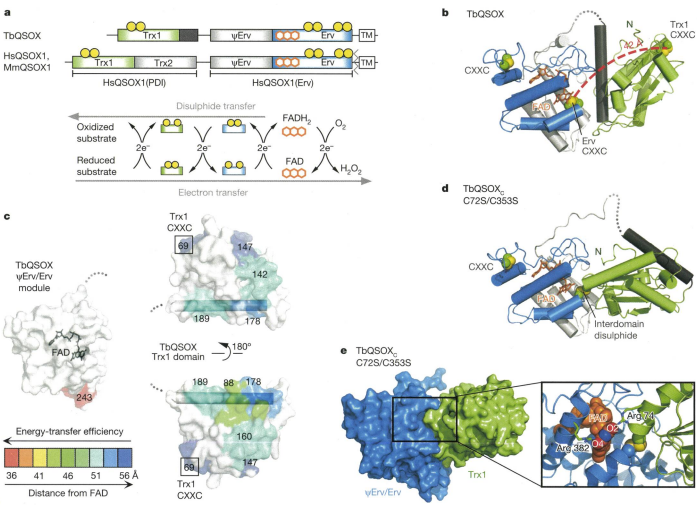


Figure 1 | TbQSOX undergoes domain reorientation to accomplish disulphide relay. a, Cys-X-X-Cys motifs are illustrated as pairs of yellow balls in maps of trypanosome and mammalian QSOX enzymes. Other cysteines are shown in Supplementary Fig. 13. Fused hexagons represent the FAD cofactor. A degenerate Erv-like domain[17] is designated 'ψErv'. Arrows depict the electron-transfer relay from reduced substrates to molecular oxygen and the corresponding outward flow of disulphide equivalents from QSOX to its substrates. TM, transmembrane. **b**, TbQSOX structure coloured according to **a**. Disulphides in Cys-X-X-Cys motifs are in space-filling representation (Cβ atom, green; sulphur, yellow). Structure stereo views are depicted in Supplementary Fig. 14. **c**, Colour-coded intramolecular distances between fluorophore-labelled TbQSOX cysteines and the bound FAD cofactor. **d**, Structure of $TbQSOX_C$. **e**, Surface representation of $TbQSOX_C$ and zoom into the interdomain interface. FAD is in space-filling representation, and oxygen atoms in the isoalloxazine ring are coloured red and labelled. Structure figures were made using PyMOL (http://www.pymol.org).

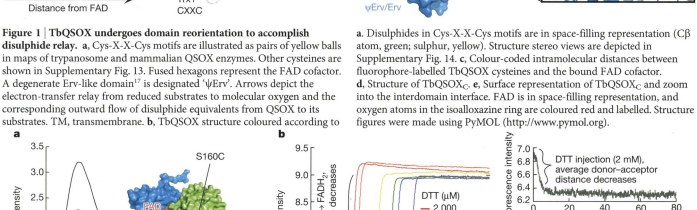


Figure 2 | Changes in the TbQSOX conformational ensemble during catalysis. a, Fluorescence of Pacific Blue conjugated at position 160 in resting TbQSOX (blue), TbQSOX oxidizing the model substrate dithiothreitol (DTT; red) and after oxygen depletion and conversion of the energy acceptor FAD to $FADH_2$ (black). Inset, surface representation of $TbQSOX_C$, as in Fig. 1e, with FAD and donor-labelling site indicated. **b**, Kinetics of donor fluorescence during TbQSOX oxidation of DTT. On DTT injection, labelled enzyme rapidly converts to a state showing greater FRET efficiency (top right panel). Other panels show longer timescales at various initial DTT concentrations. At starting DTT concentrations above ~200 μM (left panel), oxygen becomes limiting, the flavin becomes trapped as $FADH_2$ (see Fig. 1a), flavin absorbance at ~450 nm drops markedly and donor fluorescence increases correspondingly. At starting DTT concentrations of 200 μM or below (bottom right panel), DTT is limiting and the enzyme returns to its oxidized, resting state.

non-evolved interfaces (represented by crystal contacts) and less than stable protein–protein interactions[14].

In addition to illustrating the functionally important reorientation of the redox-active sites, the $TbQSOX_C$ structure shows how closure of the Trx1 domain over the FAD-binding site may enhance the active-site chemistry for disulphide formation. In the reduced, anionic form of flavin, negative charge is distributed over atoms N1, O2 and O4 of the isoalloxazine ring. Other disulphide-forming enzymes position a basic residue near the cofactor to stabilize this negative charge[15,16]. In $TbQSOX_C$, the Trx1 domain contributes an arginine side chain (Arg 74) as an interaction partner for the FAD O2 atom (Fig. 1e), which may promote formation of a thiol–flavin charge-transfer complex and, in turn, a reduced state of FAD. Indeed, the TbQSOX Arg74Ala mutant is substantially less active than wild type (Supplementary Fig. 5). Other basic residues within the Erv domain (that is, His 356 and Arg 382) may further modulate the electrostatic environment of the FAD.

A second FRET study was conducted, this time monitoring energy transfer kinetically in the presence of substrate. The TbQSOX mutant with a donor fluorophore at residue 160 (Supplementary Table IIIc) showed a reproducible FRET increase when substrate was added (Fig. 2a). As expected, when oxygen was not limiting, the FRET signal returned to resting values as the substrate was exhausted (Fig. 2b). These observations indicate that the labelled position is closer to FAD in cycling TbQSOX than in the resting, native ensemble, further supporting conformational changes during catalysis.

Although TbQSOX has conserved functional features characteristic of the QSOX family, it is one of the most divergent orthologues. TbQSOX lacks a domain (Trx2) found in metazoan QSOX enzymes (Fig. 1a), and its Trx1 redox-active motif, Cys-Gly-Ala-Cys, differs from the Cys-Gly-His-Cys common to other QSOX proteins and to PDI-family proteins. We therefore expanded our study to include the architecture and dynamics of mammalian QSOX enzymes.

The structure of *Homo sapiens* QSOX1 (HsQSOX1) was solved in two complementary fragments, $HsQSOX1(PDI)$ and $HsQSOX1(Erv)$[17] (Fig. 1a), crystallized separately (Fig. 3a and Supplementary Fig. 6). The HsQSOX1 domains containing redox-active Cys-X-X-Cys motifs were found to be similar to their TbQSOX counterparts (Supplementary

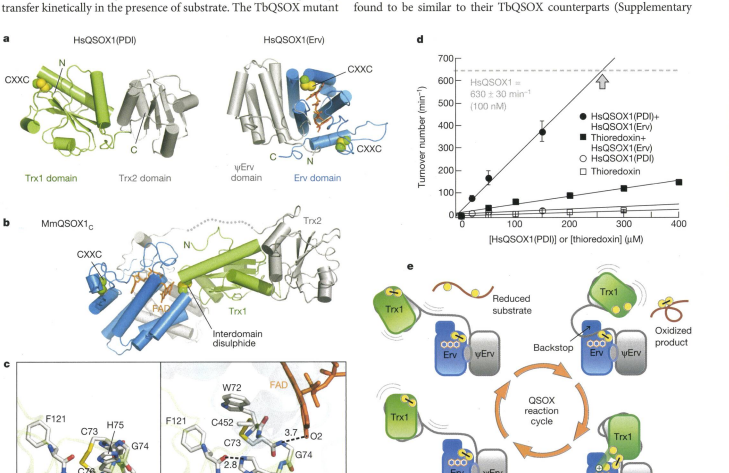


Figure 3 | Mammalian QSOX and mechanistic insights into the QSOX catalytic cycle. a, The structures of HsQSOX1(PDI) and HsQSOX1(Erv). Disulphides in Cys-X-X-Cys motifs are shown. Other cysteines appear in Supplementary Fig. 13. **b**, Structure of $MmQSOX1_C$. **c**, A comparison of HsQSOX1(PDI) (left) with $MmQSOX1_C$ (right) shows rearrangements in the Trx1 redox-active region on formation of the disulphide-transfer intermediate. Interatomic distances (dashed lines) are in Ångstrom. Residue numbering is according to MmQSOX1. **d**, Tethering increases the effective concentration of the HsQSOX1 disulphide-transferring module. Initial oxygen consumption rates were recorded for 100 nM HsQSOX1(Erv) and varying concentrations of HsQSOX1(PDI) (filled circles) or thioredoxin (filled squares) after injection of DTT. Grey dashed line indicates the turnover number for intact HsQSOX1, measured at 100 nM. By extrapolation, ~250 μM HsQSOX1(PDI) with 100 nM HsQSOX1(Erv) would be expected to support a similar reaction rate (grey arrow). Error bars, s.d. **e**, Summary of the structural basis of the QSOX catalytic mechanism, as shown in this study. Adjacent yellow balls with black bar indicate disulphide bonds; separated yellow balls indicate reduced thiols or thiolates. The '+' symbols in the closed conformation (bottom right) represent arginine side chains from either the Trx1 domain (in TbQSOX) or the Erv domain (in TbQSOX and mammalian QSOX1) that may contribute to the electron-withdrawing ability of the FAD. The 'backstop' represents conserved Erv domain loops (Supplementary Fig. 9).

Table 4). Outside of these regions, the human and parasite structures are more divergent. In particular, HsQSOX1 has the Trx2 domain but lacks the long helix of TbQSOX (Fig. 1b). The two functional modules of HsQSOX1 seem to be tethered flexibly to one another, as found for TbQSOX (Supplementary Fig. 7 and Supplementary Table 5).

Mammalian QSOX mutants designed to mimic the interdomain disulphide-transfer intermediate, analogous to $TbQSOX_C$, were constructed. The *Mus musculus* structure, $MmQSOX1_C$, solved to 2.4 Å resolution, demonstrated how disulphide transfer readily accommodates the extra domain in metazoan QSOX (Fig. 3b). Moreover, some features that may orient the two enzyme halves are conserved (Supplementary Figs 8 and 9). One difference between $TbQSOX_C$ and $MmQSOX1_C$, however, is the presence of the histidine residue (His 75) in the MmQSOX1 Trx1 Cys-Gly-His-Cys motif. His 75 appeared in two distinct rotamers in the $MmQSOX1_C$ crystals, providing insight into the dynamics of catalysis (Supplementary Figs 10 and 11). In one configuration, His 75 would be in a position to interact transiently with the thiolate of the Erv active-site cysteine Cys 452 during resolution of the interdomain disulphide intermediate (Supplementary Fig. 11). A comparison of HsQSOX1(PDI) and $MmQSOX1_C$ shows how the end of the Trx1 redox-active helix rearranges on interaction with the FAD-binding domain (Fig. 3c). No Trx1 side chain comparable to TbQSOX Arg 74 interacts with the flavin in $MmQSOX1_C$, but unpaired -NH groups at the amino terminus of the $MmQSOX1_C$ Trx1 redox-active helix approach the FAD O2 atom (Fig. 3c). Furthermore, Arg 490 is recruited from within the Erv domain for interaction with FAD O4 (Supplementary Fig. 11), like its corresponding residue in TbQSOX (Arg 382). In summary, many structural and dynamic properties of mammalian and trypanosome QSOX proteins are similar, and both enzymes seem to recruit electron-withdrawing groups to modulate active-site electrostatics during catalysis. However, the mammalian QSOX crystal structures show for the first time, to our knowledge, the plasticity and sophistication of the Cys-Gly-His-Cys motif, present in numerous redox-active proteins of the PDI family[18].

The expansion of multi-domain protein architectures contributes to the growth of the protein universe[19] and the development of complex species[20]. Rarely is it possible, however, to determine quantitatively how domain fusion contributes to protein function. Owing to its modular structure, QSOX provides an opportunity to quantify the benefit of concatenation of functional units. We measured the extent of this benefit by adding increasing concentrations of HsQSOX1(PDI) to catalytic amounts of HsQSOX1(Erv) to determine the ratio necessary to achieve the reaction rate of intact HsQSOX1 (Fig. 3d). We observed that the effective concentration of the disulphide-transferring module is increased by more than 2,500-fold by its concatenation to the disulphide-generating module through a poorly conserved linker (Supplementary Fig. 12). Our results emphasize the principle, applicable to design and *in vitro* evolution of new catalytic or signalling relays, that the function of proteins acting in tandem may be greatly improved by concatenation using flexible generic linkers.

Although mammalian cells are replete with PDI-family proteins[18], which contain diverse numbers and arrangements of redox-active and -inactive domains, QSOX is the only enzyme with an oxidoreductase domain homologous to PDI fused to a domain that generates disulphide bonds *de novo*. QSOX is also exceptional in having a primary localization outside of the endoplasmic reticulum. The QSOX structures reported here show how the two functions of the enzyme, generation of disulphides and transfer to substrate proteins, are permitted through a loose tethering of the two modules (Fig. 3e). On one hand, the fusion is sufficient to greatly increase the effective concentration of the two active sites, but on the other it allows an enormous range of motion. Understanding how this flexibility may further contribute to interactions with substrate must await identification of the native targets of QSOX catalysis in the late secretory pathway and extracellular environment.

METHODS SUMMARY

Protein preparation. TbQSOX and its mutants were produced as described[13]. Enzymes were purified using nickel-nitrilotriacetic acid (Ni-NTA), digested with thrombin and reapplied to Ni-NTA. Unbound material was collected, concentrated and purified by gel filtration. HsQSOX1(PDI) and MmQSOX1 (Cys76Ala/Cys455Ser) were produced in BL21(DE3) pLysS *E. coli* and purified as for TbQSOX. For crystallization, HsQSOX1(PDI) was treated with a dimethylamine–borane complex[21]. TbQSOX (Cys72Ser/Cys353Ser) and MmQSOX1 (Cys76Ala/Cys455Ser) were incubated before gel filtration with maleimide-functionalized polyethylene glycol (PEG) to separate $TbQSOX_C$ and $MmQSOX1_C$ from thiol-containing species.

Protein crystallization. TbQSOX crystals were grown by vapour diffusion over a well solution containing 14–18% (w/v) 2 kDa PEG monomethyl ether and 10% ethylene glycol. Crystals of $TbQSOX_C$ were grown over 20% (w/v) 3.35 kDa PEG, 0.2 M sodium sulphate and 0.5% (w/v) polyvinylpyrrolidone K15. Crystals of methylated HsQSOX1(PDI) were grown over 20% (w/v) 3.35 kDa PEG and 0.2 M lithium sulphate monohydrate. Crystals of $MmQSOX1_C$ were grown over 0.1 M MES buffer, pH 6.5, and 12% (w/v) 20 kDa PEG (orthorhombic) or 0.1 M HEPES buffer, pH 7.0, 10% (v/v) 1,5-pentanediol and 5% (v/v) glycerol (monoclinic).

Structure solution. The TbQSOX structure was determined by molecular replacement using a partial human QSOX structure (Protein Data Bank (PDB) code 3LLK) and a TbQSOX Trx1 homology model. The $TbQSOX_C$ structure was determined by molecular replacement using wild-type TbQSOX fragments. The HsQSOX1(PDI) structure was determined as described[22] (search models, PDB codes 3ED3 and 2E0Q). The $MmQSOX1_C$ structure was determined by molecular replacement using HsQSOX1(PDI) and HsQSOX1(Erv).

FRET spectroscopy. Pacific Blue C_5-maleimide labelling was according to manufacturer's instructions. The excitation wavelength was 400 nm, and energy transfer was measured with reference to enzyme containing $FADH_2$. TbQSOX (Ser160Cys) wavelength scans in the presence of dithiothreitol (1 mM) were initiated within 10 s after mixing and completed within 1 min. TbQSOX (Ser160Cys) kinetic experiments were conducted using a stopped-flow fluorimeter with a 420 nm long-pass emission filter.

Full Methods and any associated references are available in the online version of the paper.

Received 2 August 2011; accepted 28 May 2012.
Published online 15 July 2012.

1. Riemer, J., Bulleid, N. & Herrmann, J. M. Disulfide formation in the ER and mitochondria: two solutions to a common process. *Science* **324,** 1284–1287 (2009).
2. Tu, B. P. & Weissman, J. S. Oxidative protein folding in eukaryotes: mechanisms and consequences. *J. Cell Biol.* **164,** 341–346 (2004).
3. Kodali, V. K. & Thorpe, C. Oxidative protein folding and the quiescin–sulfhydryl oxidase family of flavoproteins. *Antioxid. Redox Signal.* **13,** 1217–1230 (2010).
4. Antwi, K. *et al.* Analysis of the plasma peptidome from pancreas cancer patients connects a peptide in plasma to overexpression of the parent protein in tumors. *J. Proteome Res.* **8,** 4722–4731 (2009).
5. Song, H. *et al.* Loss of *Nkx3.1* leads to the activation of discrete downstream target genes during prostate tumorigenesis. *Oncogene* **28,** 3307–3319 (2009).
6. Schulman, S., Wang, B., Li, W. & Rapoport, T. A. Vitamin K epoxide reductase prefers ER membrane-anchored thioredoxin-like redox partners. *Proc. Natl Acad. Sci. USA* **107,** 15027–15032 (2010).
7. Goodstadt, L. & Ponting, C. P. Vitamin K epoxide reductase: homology, active site and catalytic mechanism. *Trends Biochem. Sci.* **29,** 289–292 (2004).
8. Zanata, S. M. *et al.* High levels of active quiescin Q6 sulfhydryl oxidase (QSOX) are selectively present in fetal serum. *Redox Rep.* **10,** 319–323 (2005).
9. Ostrowski, M. C. & Kistler, W. S. Properties of a flavoprotein sulfhydryl oxidase from rat seminal vesicle secretion. *Biochemistry* **19,** 2639–2645 (1980).
10. Jaje, J. *et al.* A flavin-dependent sulfhydryl oxidase in bovine milk. *Biochemistry* **46,** 13031–13040 (2007).
11. Raje, S. & Thorpe, C. Inter-domain redox communication in flavoenzymes of the quiescin/sulfhydryl oxidase family: role of a thioredoxin domain in disulfide bond formation. *Biochemistry* **42,** 4560–4568 (2003).
12. Heckler, E. J., Alon, A., Fass, D. & Thorpe, C. Human quiescin-sulfhydryl oxidase, QSOX1: probing internal redox steps by mutagenesis. *Biochemistry* **47,** 4955–4963 (2008).
13. Kodali, V. K. & Thorpe, C. Quiescin sulfhydryl oxidase from *Trypanosoma brucei*: catalytic activity and mechanism of a QSOX family member with a single thioredoxin domain. *Biochemistry* **49,** 2075–2085 (2010).
14. Mitra, P. & Pal, D. New measures for estimating surface complementarity and packing at protein–protein interfaces. *FEBS Lett.* **584,** 1163–1168 (2010).
15. Inaba, K., Takahashi, Y. H., Ito, K. & Hayashi, S. Critical role of a thiolate-quinone charge transfer complex and its adduct form in *de novo* disulfide bond generation by DsbB. *Proc. Natl Acad. Sci. USA* **103,** 287–292 (2006).
16. Fass, D. The Erv family of sulfhydryl oxidases. *Biochim. Biophys. Acta* **1783,** 557–566 (2008).

17. Alon, A., Heckler, E. J., Thorpe, C. & Fass, D. QSOX contains a pseudo-dimer of functional and degenerate sulfhydryl oxidase domains. *FEBS Lett.* **584,** 1521–1525 (2010).
18. Appenzeller-Herzog, C. & Ellgaard, L. The human PDI family: versatility packed into a single fold. *Biochim. Biophys. Acta* **1783,** 535–548 (2008).
19. Levitt, M. Nature of the protein universe. *Proc. Natl Acad. Sci. USA* **106,** 11079–11084 (2009).
20. Tordai, H., Nagy, A., Farkas, K., Bányai, L. & Patthy, L. Modules, multidomain proteins and organismic complexity. *FEBS J.* **272,** 5064–5078 (2005).
21. Walter, T. S. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14,** 1617–1622 (2006).
22. DiMaio, F. *et al.* Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* **473,** 540–543 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S. Rogotner and O. Dym helped in growing the $TbQSOX_C$ crystals. We thank T. Ilani and A. Horovitz for reading of the manuscript and N. Nelson and his research group for help with X-ray data collection. A. Moseri assisted with high-performance liquid chromatography. This study was funded by the Israel Science Foundation. D.F. and A.A. acknowledge the Kimmelman Center for Macromolecular Assemblies for additional support. C.T. and V.K.K. acknowledge National Institutes of Health (NIH) grant GM26643. The molecular movie was produced using the University of California San Francisco Chimera package from the Resource for Biocomputing, Visualization and Informatics (supported by NIH grant P41 RR001081).

Author Contributions A.A. designed experiments, expressed, purified and crystallized proteins, and, together with D.F., solved and refined the TbQSOX, $TbQSOX_C$ and $MmQSOX1_C$ structures. A.A. also performed the FRET and crosslinking experiments. I.G. improved the TbQSOX crystals. Y.G. grew the $MmQSOX1_C$ crystals. V.K.K. and C.T. provided plasmids and helped to design and analyse experiments. F.D. and D.B. accomplished the molecular replacements to solve the HsQSOX1(PDI) structure. T.M. performed the mass spectrometry experiments and analyses. G.H. helped to design and analyse the FRET experiments and assisted with operation of the fluorimeter. D.F. expressed proteins, performed oxygen consumption measurements and designed and analysed the experiments. A.A. and D.F. wrote the manuscript.

Author Information Atomic coordinates and structure factors for the TbQSOX, $TbQSOX_C$ and HsQSOX1(PDI) structures have been deposited with the Protein Data Bank under accession codes 3QCP, 3QD9 and 3Q6O, respectively. The $MmQSOX1_C$ coordinates and structure factors have been deposited with accession codes 3T58 (orthorhombic) and 3T59 (monoclinic). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.F. (deborah.fass@weizmann.ac.il).

CAREERS

TURNING POINT Cardiologist is first winner of award supporting clinical research **p.421**

GRANTS Academics funded by large firms have lower publication rates **p.421**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

R. COTTAM/IMAGE200/CORBIS

TRAINING

Workshops that work

Seminars on career alternatives and soft skills can provide crucial tips for advancement. But some workshops are more helpful than others.

BY TRISHA GURA

One Thursday afternoon in May, a conference room at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, is filling beyond capacity. More than two dozen postdocs and young faculty members from the BIDMC and other affiliates of Harvard Medical School in Boston stream in for tips on how to produce an effective oral presentation. The attendees, many of whom are not native English speakers, look overworked but expectant. They have relinquished the freedom of a late-afternoon coffee break because their career advancement is on the line. They know the importance of being able to deliver an hour-long lecture or a ten-minute talk, daunting though that might be. "I have been giving talks for 30 years," begins the presenter, Terry Maratos-Flier, an endocrinologist and neurologist who directs the Office for Academic Careers and Faculty Development at the BIDMC. "So I figure I should take my expertise and offer it to you."

She does so in an hour, describing how to structure a talk, deliver it effectively and design clear, attractive slides. She engages her audience with a mixture of directives and humorous anecdotes; participants scribble notes, ask questions and offer their own accounts of success and catastrophe in public speaking.

"Sometimes I have a tendency to give more data when the audience has already got what they need," says Salin Dahlben, a psychiatrist and neurologist at the BIDMC. "I try to add something more, perhaps because of my ego."

Maratos-Flier cautions that less is more, and suggests that speakers rehearse in front of colleagues, who can offer feedback such as 'Too much information' or 'You lost me there'. By the end of the talk, many participants feel much more confident in their presentation abilities.

As 'soft' communication skills become more important in research careers, early-career scientists are increasingly attending workshops to learn them. It is no longer enough to learn by watching and consulting supervisors. "When I was in training, we could figure this stuff out on our own, stumble around or hope we could get it from our mentors," says Maratos-Flier. ▶

▶ "Now, the sense is that individuals really need to have more directed education."

Graduate students and postdocs, under pressure to publish and seek funding, have little time to learn non-research skills to boost their competitiveness. But employers — especially in industry — want recruits to have those skills. "If you have waited until year five of your postdoc and you think you are going to break into industry as a 'plan B,'" says Sibby Anderson-Thompkins, director of postdoctoral affairs at the University of North Carolina at Chapel Hill (UNC), "it may not happen."

So universities around the world are setting up schemes to teach transferable and soft skills. These might involve hour-long lectures, half-day seminars or hands-on, activity-based experiences that offer aspiring researchers advice and training on how to run a lab, create a budget, write a paper, collaborate with colleagues and, of course, deliver a good talk. But not all workshops provide useful or practical advice that is worth a fledgling scientist's time.

INFORMED DECISIONS

Early-career researchers looking for workshops should start on campus. Cathee Johnson Phillips, executive director of the US National Postdoctoral Association (NPA) in Washington DC, says that soft-skills training is now available at more than 135 offices or associations of postdoctoral affairs in the United States, as well as hundreds of career-development offices at universities and research institutions in the United States and Europe. The NPA itself organizes an annual meeting with workshops on a range of topics: everything from conflict resolution to English for postdocs. One seminar at this year's meeting, called 'Business Dining is Not About Food', taught etiquette for business lunches and dinners.

Once they have picked a provider, researchers must choose which workshops to attend. "It depends on what you want to get out of it," says Jill Hurst, a pharmaceutical and biochemical sciences postdoc at the UNC. "For students and younger postdocs trying to figure out which way they are going to go career-wise, a more formal, this-is-what-is-available panel discussion works well." For in-depth inquiries, it might be preferable to attend a lunch seminar with an industry representative, for example, says Hurst. The seminar format gives researchers a chance to ask probing questions and improve their sense of what day-to-day working lives are like in industry.

Some institutions offer more extensive experiences. Stephen Fuchs, a biochemist at Tufts University in Medford, Massachusetts, attended a leadership training course at Cold Spring Harbor Laboratory in New York. There, he and a select group spent three days listening, discussing and role-playing. "It was intense and focused," Fuchs recalls. "And everybody there was looking for the same types of things." Those included tips on negotiating salary and position, managing personnel and dealing with difficult conversations in the lab.

The specificity made it the best training experience that Fuchs has ever had, he says. To assess just how focused a workshop will be, prospective attendees can contact organizers and find out what motivated them to put the workshop together. For example, if a company representative is giving a talk with the intention of recruiting scientists for certain positions, the event might not be a great choice for a first-year postdoc who simply wants to know more about working in industry. A panel discussion by industry experts would be a better fit.

Some misses can't be predicted. While still a postdoc at the UNC, Fuchs attended a seminar on the responsible conduct of research at the US National Institutes of Health (NIH) in Bethesda, Maryland. Speakers lectured on issues of authorship, data fabrication and conflict of interest. The seminar was mandatory for researchers who wanted to obtain federal funding, but it felt too "philosophical" to Fuchs, more a reflection on ethics than a guide to being a responsible researcher.

Scientists need practical guidance, says Fuchs. So he organized his own workshop on the subject at the UNC, and made sure that it included substantive advice. He brought in experts who deal with the issues on a daily basis, such as a representative from the UNC Office of Technology Transfer to talk about intellectual-property issues, and the university's ombudsman to discuss how he investigates complaints and mediates settlements. Participants gave positive feedback, and the workshop is now held regularly.

"It is amazing, the lack of self-awareness I see."
Sibby Anderson-Thompkins

Early-career scientists should keep an eye out for active rather than passive events, so that they can practise the skills of interest rather than merely discussing them. Dara Wilson-Grant, associate director of postdoctoral affairs at the UNC, held a workshop on networking, tied to an actual campus networking event with industry representatives. Workshop participants first listened to tips on networking, then role-played with each other and finally tried their luck at the event. Hurst attended and found the workshop so helpful that she landed a job interview that led to an offer after chatting with a science writer from a local biotechnology firm. "I learned how to ask good questions and get a sense of what the company was like," she says. "A year ago, I wouldn't have been able to do that."

Before attending a workshop, scientists can research the speaker's expertise and style to ensure that they are a good fit. Dahlben's favourite thing about Maratos-Flier's workshop was her demeanour. By presenting herself as a seasoned researcher who had experienced the same travails as the audience and overcome them, Maratos-Flier had credibility with her audience, says Dahlben.

POCKETS OF RESISTANCE

Despite the need and opportunities for soft-skills training, not everyone is enthusiastic. Some see workshops as irrelevant or unnecessary. "In terms of giving a talk, I have participated in many conferences," says Bernard Charlier, a geosciences postdoc at the Massachusetts Institute of Technology in Cambridge. "I have seen very good presentations and very, very bad. So, I learn from those."

But it may not be possible to learn enough by imitation in today's competitive environment. Critiques and feedback are crucial. "It is amazing, the lack of self-awareness I see," says Anderson-Thompkins. "People think, 'This isn't something I need a lecture on.' But I have seen it in action, with networking, for example. Postdocs clear a room because of the way that they are approaching people."

Some early-career scientists might want to attend workshops, but run into resistance from supervisors who prefer their charges to stay in the lab rather than investigate alternative careers. However, workshop organizers such as Paule Biaudet, a specialist in doctoral studies and employability at Pierre and Marie Curie University in Paris, argue that professional development helps not only students and postdocs, but also their principal investigators, who should support their lab members' training aspirations. She explains to mentors that the reputation of a lab is often linked to the career successes of the people who have been trained there. "Then the supervisors feel they must let the students go to the training," says Biaudet.

Good training ultimately opens doors in an increasingly challenging job market. Maratos-Flier says that when she started out in research, "it would have been nice to have had somebody tell me in a non-judgemental way, 'you're doing this or that wrong', or a better or more effective way to do it. Now, you don't need to reinvent the wheel. Workshops can get you to a skill level you want to be at, without having to work it out all by yourself." ■

Trisha Gura *is a freelance writer based in Boston, Massachusetts.*

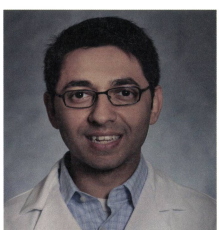
UNCJJ PRUITT

CORRECTION

The map in the story 'Excellence revisited' (*Nature* **487,** 519–521; 2012) omitted the symbol to show that the University of Konstanz has received an award for 'future concepts'. The map has been corrected online at go.nature.com/rciac4.

TURNING POINT
Nehal Mehta

Nehal Mehta is the inaugural Lasker Clinical Research Scholar — the first recipient of an award established by the US National Institutes of Health (NIH) and the Lasker Foundation in New York. Last month, Mehta moved from the University of Pennsylvania (UPenn) in Philadelphia to the NIH campus in Bethesda, Maryland, where he is studying links between cardiovascular disease and the chronic inflammation associated with conditions such as diabetes, obesity and psoriasis. The award affords a serious career boost: after 5–7 years with positive reviews, Mehta will have the option of getting tenure at the NIH or applying — with funding in tow — to an academic institution.



You trained as a medical doctor. When did you realize that you also wanted to do research?

It started with my internal-medicine residency. I had a lot of unanswered questions — for instance, why do certain people have heart attacks when they have atherosclerosis (hardening of the arteries) and others don't? My main mentor at UPenn encouraged me to investigate such questions. I enrolled about 80 people in his clinical study asking how inflammation affects molecular pathways that lead to atherosclerosis. He opened my eyes to clinical research.

What benefits does the Lasker award bring?

It provides the resources to develop my career — not just money, but also crucial elements such as core facilities, access to patients and travel opportunities. The scholarship can last up to 12 years; clinical research is very time consuming, mostly because of clinical trials. This award comes once in a lifetime: I have the chance to run my own programme and see patients at the NIH Clinical Center, a world-class research hospital, without the barriers that face many young investigators, such as the need to secure funding.

What is your research focus?

My lab is poised to answer whether inflammation is simply a by-product of certain conditions, or a driver of cardiovascular disease. One of our studies asks whether treating chronic inflammation improves markers of cardiovascular health, such as how the body handles cholesterol and sugar.

What are your plans for the NIH?

I have a three-pronged approach. We are starting a national effort to obtain skin, fat and blood samples from people with psoriasis to understand the dysregulation of molecular and cellular pathways related to cardiovascular and metabolic diseases. The second arm will bring patients into the NIH Clinical Center to study risk factors for cardiometabolic disease and assess vascular inflammation. Finally, we will examine cardiometabolic pathways in animals and cell culture.

You plan to continue working at UPenn. How will you balance this with your job at the NIH?

I will be an adjunct faculty member at UPenn, and will go there every seven weeks to see people with psoriasis in my preventive cardiology programme. Hopefully most of them will come to the NIH; we will continue to enrol patients at UPenn but do assays and imaging analysis at the NIH. I have roots at UPenn: mentors are now collaborators. And I have an excellent nurse.

Did you ever consider getting a PhD?

I have thought about it, but I don't think it is absolutely necessary. There are some things I struggle with in the lab, such as operating certain equipment, but I can solve those problems by surrounding myself with great scientists. I do believe that a certain amount of advanced training is necessary. I have a master's degree in clinical epidemiology, which helps me to do high-level clinical research, design studies and work in a lab.

How do you juggle medicine and research?

My patients provide me with rich samples and rich questions. I get a lot out of each activity: by seeing patients, I am helping people but also doing research. ■

INTERVIEW BY CHARLOTTE SCHUBERT

GRANTS

Industrial outcomes

Academic scientists funded by large companies publish, patent and are cited less often than those supported by smaller firms or other sources, a study says. *Research Grants, Sources of Ideas and the Effects on Academic Research*, released on 29 July by the Centre for European Economic Research in Mannheim, Germany, polled researchers at 46 German universities. Co-author Cornelia Lawson, an economist at the University of Turin, Italy, says that large grant-giving firms might modify the focus of the work during the project so that there are not enough data to publish after the grant. They also tend to fund applied research, which may be cited less often, she says. The study notes that industrial sponsorship has been on the rise globally for several years.

EQUALITY

Awards for women

Five early-career female researchers from Africa, the Middle East, Asia, Latin America and the Caribbean are to be honoured with US$5,000 Elsevier Foundation Awards in a programme to encourage women to pursue science in nations that lack scientific expertise, resources and gender equality. "They are not given the opportunity to do good science," says Peter McGrath, a programme officer for co-sponsor TWAS, the academy of sciences for the developing world in Trieste, Italy. "This is a way to bring their work to the international forefront." Nominations for researchers from 81 eligible nations who earned PhDs in the past 10 years will be accepted until 30 September.

GERMANY

Initiative criticized

Germany's €4.6-billion (US$5.6-billion) Excellence Initiative has not helped universities to distinguish themselves, according to a briefing (go.nature.com/nsjpca) released on 22 June by the Social Science Research Center Berlin (WZB). The report says that institutions are concentrating on basic research to the detriment of teaching and applied research. "When we talk about excellence, we also have to talk about conditions for excellent teaching," says co-author Dagmar Simon, head of science-policy studies at the WZB. The scheme to make Germany's university system more competitive is in its second phase (see *Nature* **487**, 519–521; 2012).

SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Simon Fraser University is one of Canada's leading comprehensive universities with three campuses in Metro Vancouver. Through the strength and commitment of our staff and faculty, SFU is recognized as a top employer in both BC and Canada. We invite applications for the following position:

DEPARTMENT OF CHEMISTRY
FACULTY POSITION IN CHEMISTRY

The Department of Chemistry at Simon Fraser University (SFU) invites applications for a tenure-track Assistant Professor position in Bio-Physical or Materials Chemistry to take effect in September 2013, subject to final budgetary approval.

Applicants should have a Ph.D. degree and will normally have postdoctoral or industrial experience. Outstanding candidates with a commitment to excellence in research and teaching are being sought. The candidates will be expected to develop and maintain both an innovative, externally funded research program, and an excellent teaching record at both the undergraduate and graduate levels in Chemistry.

All qualified candidates are encouraged to apply; however, Canadians and Permanent Residents will be given priority. Simon Fraser University is committed to an equity employment program that includes special measures to achieve diversity among its faculty and staff and welcomes applications from all qualified individuals. We therefore particularly encourage applications from qualified women, aboriginal Canadians, persons with disabilities, and members of visible minorities. Applicants should send a complete résumé, a concise research proposal, a short teaching dossier, and a list of three individuals willing to act as referees with their addresses, telephone and/or fax numbers, and e-mail addresses. All correspondence should be sent to:

Dr. Zuo-Guang Ye
Professor and Chair
Department of Chemistry
Simon Fraser University
8888 University Drive
Burnaby BC V5A 1S6 Canada
E-mail: chemchr@sfu.ca

The competition will remain open until the position is filled. Screening of applications will commence on November 1, 2012.

Please note: We thank all applicants for their interest, however, only those selected for interviews will be contacted.

Please note that under the University Act, personal information that is required by the University for academic appointment competitions will be collected. For further details see the Collection Notice. (www.chemistry.sfu.ca/about/employment)

NW227378R

Freie Universität Berlin

Department of Biology, Chemistry and Pharmacy - Institute of Chemistry and Biochemistry
invites applications for a tenured

Professorship in Biochemistry with focus on RNA Biochemistry (W 2)

The successful applicant will be required to teach courses in Biochemistry and focus on RNA Biochemistry in her or his research.
Appointment requirements are governed by article 100 of the Berlin Higher Education Act (Berliner Hochschulgesetz).
The appointee will have an outstanding track record of university teaching and research with a strong international component. She or he will be an aspiring scientist with an international reputation and a research interest in the molecular mechanisms of processes based on or regulated by RNA or ribonucleoproteins. The range of methods to be employed should complement the expertise at the institute and enable collaborations with other members of the department. Candidates should have experience acquiring funding for, and carrying out, sponsored research as well as designing and implementing curricula in biochemistry.
The newly appointed professor will be required to contribute to the department's long-term efforts to establish research initiatives in RNA Biochemistry. A possibility to embed her or his research activity within existing networks such as the Collaborative Research Centers Scaffolding of Membranes - Molecular Mechanisms and Cellular Functions (SFB 958) or Multivalency as Chemical Organization and Action Principle (SFB 765) as well as the Focus Area Nanoscale is desirable but not mandatory.
Applications quoting the **reference code W2-Prof.-RNA-Biochemie** and including, apart from standard documents (CV, publications, diplomas and certificates), a list of previously acquired third-party funds (including the amounts raised) and previous courses taught, a description of planned future research and potential partners in Berlin (2 pages max.) as well as a brief overview and justification of rooms, equipment and personnel needed must be received in hard copy and on a CD-ROM no later than **September 13th, 2012** at Freie Universität Berlin, Fachbereich Biologie, Chemie, Pharmazie, Dekan Herrn Prof. Dr. Hans-Ulrich Reißig, Takustr. 3, 14195 Berlin (Dahlem, Germany).
Application guidelines and general information on the appointment procedure as well as requirements for professorships at Freie Universität Berlin can be found at www.fu-berlin.de/praesidialamt. For additional details, please visit www.fu-berlin.de, www.bcp.fu-berlin.de, www.sfb765.de, www.sfb958.de and www.nanoscale.fu-berlin.de.

W227403R

---

university of groningen

The Faculty of Mathematics and Natural Sciences has a vacancy for a **Full Professor in Theoretical Chemistry** in the Zernike Institute for Advanced Materials.

We look for an enthusiastic theoretical chemist who is an outstanding teacher and performs internationally leading research in the development and application of the theory for molecular systems. Close collaboration with partners within and outside the Zernike Institute for Advanced Materials is expected.

For the complete description of the vacancy and additional information about the application procedure, please visit our website: www.rug.nl

W227032R

Follow *Naturejobs* on Twitter

For the latest career news: **twitter.com/naturejobs**
For all job listings: **twitter.com/naturejobs_feed**

twitter

---

naturejobs

Read our Postdoc articles and news in the **News and Comment** section of **naturejobs.com**

- Helpful advice
- Real life experiences
- Follow the highs and lows of postdoc life

naturejobs

European Head Office, London naturejobs@nature.com
The Macmillan Building, 4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961 Fax: +44 (0) 20 7843 4996

Global Head:
Nils Moeller; +44 (0) 20 7843 4953 n.moeller@nature.com

Southwest UK/RoW:
Emma Hare; +44 (0) 20 7843 4944 e.hare@nature.com

Northeast UK/Ireland:
Matthew Ward, +44 (0) 20 7014 4059 m.ward@nature.com

France/Switzerland/Belgium:
Julie Aramburu; +44 (0) 20 7843 4994 j.aramburu@nature.com

Scandinavia/Spain/Portugal/Italy:
Megan Ball, +44 (0) 20 7843 4973 m.ball@nature.com

North Germany/The Netherlands/Eastern Europe:
Kristin Doelling; +44 (0) 20 7843 4970 k.doelling@nature.com

South Germany/Austria:
Maria Kubalova, +44 (0) 20 7014 4084 m.kubalova@nature.com

Natureevents:
Matthew Clare, +44 (0) 20 7014 4015 m.clare@nature.com

Business Development Manager, Middle East:
Jon Giuliani, +44 (0) 20 7843 4922 j.giuliani@nature.com

USA Head Office, New York naturejobs@us.nature.com
75 Varick Street, 9th Floor, New York, NY 10013-1917
Tel: +1 800 989 7718 Fax: +1 800 989 7103

US Sales Manager:
Ken Finnegan, +1 212 726 9248 k.finnegan@us.nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2–37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751, Fax: +81 3 3267 8746

Asia-Pacific Sales & Business Development Manager:
Yuki Fujiwara, +81-3-3267-8765 y.fujiwara@natureasia.com

Business Development Manager, Greater China/Singapore:
Gloria To, +852 2811 7191 g.to@natureasia.com

India
3a, 4th Floor, DLF Corporate Park, Gurgaon, 122 002, India

Advertisement Sales:
Vikas Chawla, +91 1242881057 v.chawla@nature.com

POSTDOCTORAL ASSOCIATES

The Duke Human Vaccine Institute and the Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID), providing national and international leadership in the fight against major infectious diseases, is currently recruiting for Postdoctoral Associates.

Duke University and the Duke Human Vaccine Institute, under the direction of Dr. Barton Haynes, have been awarded a new grant, the Duke Center for HIV/AIDS Vaccine Immunology-Immunogen Discovery grant, to overcome roadblocks that hinder the development of a vaccine for HIV, and to design and develop immunogens for a successful vaccine. As a result of this seven-year, $139-million grant, the Duke Human Vaccine Institute is currently recruiting for postdoctoral associates to participate in this work.

Projects will include study of the B cell response to HIV envelope proteins in non-human primates and in human vaccinees, study of B cell regulation in humanized mice, study of B cell development in rhesus macaques and humans, and studies of the role of factors that regulate T follicular helper cell generation by immunogens. In addition, the regulation of somatic hypermutation and class switching of immunoglobulins will be of prime interest.

The education, expertise, initiative and dedication of the postdoctoral associates engaged in this project will be important to the success of the studies. This fellowship will be an opportunity to become part of a winning team that is working to discover novel ways to prevent HIV infection and, therefore, protect people world-wide. Qualified candidates must have a Ph.D. in virology, immunology, molecular biology, computational biology or related disciplines.

Candidates should send a cover letter and current curriculum vitae to:

Duke Human Vaccine Institute
Email: dhvi.careers@nates.duke.edu
(Please reference "PD-DHVI" in subject line of email)

Duke Human Vaccine Institute

CHAVI-ID
Duke Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery

Duke University is an Equal Opportunity/Affirmative Action Employer

NW227351RL

Discover the next breakthrough. Challenge the future of science.

BE PART OF THE FOUNDING FACULTY OF SCIENTISTS AND ENGINEERS

The Singapore University of Technology and Design (SUTD), established in collaboration with the Massachusetts Institute of Technology (MIT), is seeking exceptional faculty members to lead its first batch of students.

SUTD, the first university in the world with a focus on design accomplished through an integrated multi-disciplinary curriculum, has a mission to advance knowledge and nurture technically grounded leaders and innovators to serve societal needs. Design, as an academic discipline, cuts across the curriculum and is the framework for research and educational programmes.

MIT's collaboration with SUTD includes the development of new courses and curricula, assistance with the early deployment of courses in Singapore, faculty and student recruiting, student exchanges, mentoring, career development, and collaborating on major joint projects through a new international design centre. Newly hired SUTD faculty may spend up to a year at MIT in a specially tailored programme for collaboration and professional development.

FACULTY POSITIONS

The qualifications for the faculty position include: an earned doctorate in Biology, Biomedical Sciences, Chemistry, Chemical Engineering, Physics and Mathematics, a strong commitment to teaching at the undergraduate and graduate levels, a demonstrated record of, or potential for scholarly research, and excellent communication skills.

We invite applications for interdisciplinary faculty appointment at all levels, with many opportunities available in particular at the Assistant and Associate Professor levels. Duties include teaching of graduate and undergraduate students, research, supervision of student research, advising undergraduate student projects, and service to SUTD and the community. Faculty will be expected to develop and sustain a strong research programme. Attractive research grant opportunities are also available.

Successful candidates can look forward to internationally competitive remuneration, and assistance for relocation to Singapore. If you want to be part of the founding faculty that will shape the world of tomorrow, please apply to SUTD at **www.sutd.edu.sg**

SUTD
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
Established in collaboration with MIT

JP227316R

Newcastle University

wellcome trust Strategic Award

Wellcome Trust Centre for Mitochondrial Research

A new Wellcome Trust Centre for Mitochondrial Research has been established at Newcastle University. The aim of the Centre is to better understand the cause and develop methods to treat or prevent mitochondrial disease. There are four main research themes that are designed to synergistically develop both basic and clinical science. For each of these themes there are positions available:

- Mitochondrial gene expression (2 Postdoctoral Scientists, 1 Research Assistant/ Associate (Scientific Support), 1 PhD Student)
- Development of IVF based techniques to prevent mitochondrial disease (1 Postdoctoral Scientist, 1 Research Assistant/Associate (Scientific Support), 1 PhD Student)
- The neurology of mitochondrial disease (1 Clinical Research Associate, 1 Research Assistant/Associate (Scientific Support))
- The role of mitochondrial DNA in common complex diseases (1 Bioinformatician, 1 Statistician, 1 Research Assistant/Associate (Scientific Support), 1 PhD student)

For further information please visit
http://www.ncl.ac.uk/iah/research/centres/wellcome/
For vacancies please see
http://www.ncl.ac.uk/vacancies/jobs/ and
http://www.ncl.ac.uk/iah/education/studentship.htm

Closing date: 14 September 2012.

U227644R

Committed to Equal Opportunities

www.ncl.ac.uk/vacancies

European Environment Agency

Call for Expressions of Interest for experts to be appointed members of the Scientific Committee of the European Environment Agency

EEA/SC/2012/001-010

Scope of the call

The EEA organises this call to appoint approximately 5 Scientific Committee members with the expertise in one or more of the following priority areas as identified by the EEA Management Board:

Biodiversity, nature conservation and ecosystems management; Terrestrial ecosystem and soils; Marine ecosystems and maritime space; Urban systems and quality of life; Systems resilience and adaption to environmental change; Chemicals and the environment; Environment policy; Environment statistics and accounting; Environmental information systems; Environmental monitoring.

Interested candidates are invited to look at the full vacancy notice published on the EEA website http://www.eea.europa.eu/about-us/governance/scientific-committee/call-for-expression-of-interest

Deadline for applications:
Wednesday 12 September, 12:00 CET

W227607R

UNIVERSITY OF CAMBRIDGE

A world of opportunities
www.cam.ac.uk/jobs/

BioRepository Informatics Manager

Institute of Public Health
£27,578 - £35,938 pa

An exciting new opportunity has arisen for a BioRepository Informatics Manager to be responsible for the day to day laboratory informatics in the brand new NIHR BRC-MRC BioRepository on the Cambridge Biomedical Campus at Addenbrooke's Hospital. The main role of the post-holder will be to provide laboratory informatics infrastructure and research support for studies. The research will involve a range of activities including genetic research, sample management, high-throughput biomarker related work and epidemiological studies. Candidates should be qualified to at least an honours degree level in computer science or a related discipline. The ideal candidate will have experience in configuring, developing and maintaining a Laboratory Information Management System (LIMS) and have experience of managing laboratory data within human research studies. Limit of tenure: four years in the first instance.

Further Particulars and details of how to apply can be accessed from www.phpc.cam.ac.uk/jobs

Informal enquiries regarding this position can be made to Matt Sims (matt.sims@mrc-epid.cam.ac.uk).

Quote reference: SY00486, Closing date: 10 September 2012 at 5pm. Interview date: 17 September 2012.

The University is committed to Equality of Opportunity.

U227643R

UNIVERSITY OF LIVERPOOL

Faculty of Health and Life Sciences
Institute of Integrative Biology

Postdoctoral Research Associate

£31,020 - £35,939 pa

We are seeking an experienced individual for a BBSRC funded project to identify the mechanisms through which the tsetse fly symbiont *Sodalis* alters tsetse susceptibility to trypanosomes. You should have experience of using cell culture, molecular biology and high throughput sequencing to generate genomic and RNASeq data. The project involves the Centre for Genomic Research, a leading laboratory for the application of genomic technologies and bioinformatic analyses. A PhD in a biology related subject is essential. The post is available for 3 years.

Job Ref: R-580285/N **Closing Date: 10 September 2012**

For full details, or to request an application pack, visit www.liv.ac.uk/working/job_vacancies/ or e-mail jobs@liv.ac.uk Please quote job ref in all enquiries.

COMMITTED TO DIVERSITY AND EQUALITY OF OPPORTUNITY

Stonewall DIVERSITY CHAMPION

U227631R

www.naturejobs.com

Department of Health and Human Services
National Institutes of Health
National Institute on Minority Health and Health Disparities

The Division of Intramural Research (DIR) of the National Institute on Minority Health and Health Disparities (NIMHD) is seeking an exceptional candidate for the position of Deputy Scientific Director/Senior Investigator to provide leadership and support as an active partner with the Scientific Director in leading the research program. The inter-multidisciplinary intramural research program under development has three pillars: basic, clinical/translational and basic behavioral/social research. This position offers a unique and exciting opportunity for the right individual to share responsibility in providing visionary leadership to an organization dedicated to developing the science that will lead to the elimination of health disparities. A candidate is sought who has a broad view of research and the personality to build partnerships with intramural and extramural scientific communities. The incumbent will serve as the liaison between the DIR and the NIMHD Board of Scientific Counselors (BSC) with full oversight responsibilities for the entire BSC process. The incumbent will also build trans-NIH scientific and clinical collaborations and participate in trans-NIH initiatives. The successful candidate will also be provided independent resources to develop his/her own research program as and will be eligible for tenure as a senior investigator within DIR, NIMHD.

QUALIFICATIONS: Applicants must have an **M.D., Ph.D.**, or both as well as an active, top quality independent research program in one or more scientific areas, related to the above mentioned DIR areas of interest. The candidate shall have administrative experience running a complex research program. The candidate should be a strong communicator with the ability to work collaboratively to solve problems and to make informed decisions.

SALARY/BENEFITS: The Deputy Scientific Director, NIMHD will be appointed at a salary commensurate with his/her qualifications and experience. Full Federal benefits, including leave, health and life insurance, retirement and savings plan (401K equivalent) will be provided. This position is subject to public financial disclosure requirements.

HOW TO APPLY: Interested candidates should send a letter of interest, including a statement of research and administrative experience; curriculum vitae and bibliography; statement of research interests, and three letters of reference. Application packages should be sent to the National Institutes of Health (NIH), National Institute on Minority Health and Health Disparities, **c/o John O'Shea, Ph.D., Chair, Search Committee, 6707 Democracy Blvd., Suite 800, Bethesda, Maryland 20892 or electronically: nimhdirp@mail.nih.gov; Subject Line: NIMHD Deputy Scientific Director Search.** For further information about the position, please contact Ms. Keta Lawson at 301-496-1366 or nimhdirp@mail.nih.gov.

All information provided by the candidates will remain confidential and will not be released outside the NIMHD search process without a signed release from candidate.

The NIH encourages the application and nomination of qualified women, minorities and individuals with disabilities.

APPLICATION MUST BE RECEIVED BY FRIDAY, OCTOBER 5, 2012

DHHS and NIH are Equal Opportunity Employers

NW227462R

Chief Scientist to lead a new laboratory
in Nuclear Physics, RIKEN, JAPAN

RIKEN invites applications for the position of Chief Scientist (Laboratory Director) to lead a new laboratory in experimental nuclear physics using the RI Beam Factory (RIBF) of RIKEN Nishina Center for Accelerator-Based Science.

The present position is a permanent appointment, subject to RIKEN's mandatory retirement age of 60.
RIKEN expects that the successful applicant will be able to take up this position on April 1st, 2013.

Applicants should send a full curriculum vitae and photograph; list of publications; one copy each of five key publications; a statement (about five pages A4 sized paper) explaining former research experience, and proposals for research at RIKEN; and the names and addresses of three referees.

All applications should reach RIKEN by September 24th, 2012.

Applicants should address all correspondence to:
Dr. Hiroyoshi Sakurai, Chair of the Chief Scientist Search Committee, RIKEN Nishina Center for Accelerator-Based Science 2-1 Hirosawa, Wako, Saitama 351-0198, JAPAN
E-mail address: sakurai@ribf.riken.jp

For more information, please visit:
http://www.riken.jp/engn/r-world/info/recruit/k120924_e_rnc.html

RIKEN

JP227075R

5 Tenure-Track Positions
Hokkaido University 2012 (Sapporo, Japan)

Hokkaido University invites applicants for Tenure-Track Assistant Professor Level Positions with **starting as soon as possible after December 16,2012.**

We seek young researchers who have all the qualities that are necessary to become next generation leaders in basic interdisciplinary research areas.

■ Positions: **Five Tenure-Track Assistant Professors**
24-4 :Graduate School of Medicine
24-5 :Faculty of Agriculture
24-6 :Faculty of Engineering
24-7 :Faculty of Science (Chemistry)
24-8 :Faculty of Science (Biological Sciences)
Please see
Job Opening in the Leader Development Station (L-Station)
homepage for more information.
http://www.cris.hokudai.ac.jp/tenure-track/

■ Criteria:
Applicant must have a doctorate obtained no longer than 10 years or with equivalent research experience (including PhD acquired from a foreign university)
Applicant must be under 40 years old (project policy of MEXT)

■ Application Deadline: **MUST ARRIVE by 5pm October 9, 2012, Japan time.**
*We accept applications by postal mail only.

■ **Research fields on the homepage:**
http://www.cris.hokudai.ac.jp/tenure-track/

■ Leader Development Station (L-Station)
Research Department
Hokkaido University Creative Research Institution
North 21 West 10, North ward,
Sapporo 001-0021, JAPAN
TEL: +81-11-706-9181, 9283 FAX: +81-11-706-9272
Email: **lstation@cris.hokudai.ac.jp**

*Hokkaido University operates the "Hokkaido University Leader Development System in the Basic Interdisciplinary Research Areas" project, under FY 2012 project "Tenure-Track System Promotion Program" funded by Ministry of Education, Culture, Sports, Science and Technology(MEXT).

JP227428R

Cancer Research Initiatives Foundation (CARIF)

Cancer Research Initiatives Foundation, which opened in 2001, is the only non-profit cancer research laboratory dedicated to Malaysian cancer research. Over the years, our scientists, in collaboration with researchers both locally and worldwide, have achieved advances in the understanding of cancers, with a particular emphasis on cancers which are prevalent locally and regionally and using resources available from Malaysia's biodiversity [www.carif.com.my].

CARIF is committed to developing a workforce of the highest caliber. We aim to recruit dedicated and determined individuals whose interest in working with us is not just driven by a desire to do something interesting or worthwhile, but by a sincere commitment to developing results as part of an innovative, learning environment.

Research Group Leaders. CARIF is expanding and looking to recruit 3 to 5 outstanding scientists to head research groups, with priority given to bioinformatics, nasopharyngeal cancer, drug discovery and development, and immunotherapeutics. The successful Research Group Leader candidate must have a stellar track record of research accomplishments (e.g. independent funding and relevant publications) and an enthusiasm for building new research initiatives. All Research Group Leaders are expected to initiate and sustain vigorous research programmes that both advance and are relevant to CARIF's mission. All Research Group Leaders will be provided with additional personnel supported from core funding. Salary will commensurate with qualifications and experience and starting date will be as soon as possible after the closing date [15 October 2012].

Genetic Counsellor. The Genetic Counsellor will work with a multi-disciplinary team offer pre- and post-genetic testing genetic counseling to Malaysian families and individuals with inherited predisposition to breast and ovarian cancer.

Qualified applicants are invited to provide a cover letter, curriculum vitae that details research experience and interests, and names and contact details of four referees to info@carif.com.my.

JP227402R

Drug Discovery Researcher

Faculty of Pharmacy and Pharmaceutical Sciences
Drug Discovery Biology

Remuneration: $80,431 – $88,218 pa HEW Level 07 (includes 9% employer superannuation)

Duration: 12-month fixed-term appointment

Enquiries: Dr Christopher Langmead, Head, Servier Program in Drug Discovery, 03 9903 9096

Location: Parkville campus **Job no:** 503376

You can learn more about this position and apply online at monash.edu/jobs

An Equal Opportunity Employer

MONASH University

JP227599R

You don't have to look this hard to find a job with Naturejobs.
nature jobs
making science work

Deutsches Krebsforschungszentrum, Heidelberg, Germany offers
10 Postdoctoral Fellowships in Cancer Research

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

The German Cancer Research Center (DKFZ) invites applications for 10 two-year Postdoctoral Fellowships in Cancer Research. This Program provides interdisciplinary training and research opportunities in state-of-the-art research facilities for excellent young scientists, no longer than 2 years past graduation, who wish to work at the forefront of cancer research in any of the following topics:

Cell and Tumor Biology, Developmental and Stem Cell Biology

—

Functional and Structural Genomics, Systems Biology

—

Epidemiology, Cancer Risk Factors, Cancer Epigenomics

—

Tumor Immunology

—

Infection and Cancer, Tumor Virology

—

Imaging and Radiology, Medical Physics

—

Translational Cancer Research

The DKFZ is Germany's leading Cancer Research Center situated in the biomedical campus of the Heidelberg University in close vicinity to Heidelberg's city center. Heidelberg is home of other top research centers including the European Molecular Biology Laboratory (EMBL), Max-Planck Institute for Medical Research (MPI) and the Center for Molecular Biology in Heidelberg (ZMBH), with whom the DKFZ maintains close relationships.

For applications a qualification equivalent to a PhD or MD (Dr. rer. nat., Dr. med.) is requested from all applicants. Applicants are selected in a competitive process. The application material (including a one page statement of research interests, curriculum vitae, copy of latest degree certificate and two letters of recommendation) needs to be submitted in English. In addition a list of three potential research groups/divisions in which the applicant might want to work should be submitted.

—

Submit your complete application as a PDF file to postdoc@dkfz.de
Further Info: www.dkfz.de
Deadline: September 15th, 2012

WG2736WR

Faculty Positions in Continental Margins and Coastal Environments

Scripps Institution of Oceanography (SIO) at UC San Diego (http://sio.ucsd.edu) is committed to academic excellence and diversity within the faculty, staff, and student body. SIO is a world renowned center of marine research with approximately 200 principal investigators leading research programs on all aspects of earth, ocean and atmospheric sciences.

Coastal Processes #10-439:
We seek outstanding candidates with complementary interests in the physical oceanography of the coastal environment. Topics of interest are focused on the shelf, inner shelf, and near shore, including tidal inlets, estuaries, river mouths, beaches, and headlands. Some examples of physical interest are: surf zone processes; surf zone-inner shelf exchange; sediment, chemical, and biological transport; waves (internal and surface); and coastal response to sea level changes or human activity.

Coastal Ecology #10-440:
We seek outstanding candidates with expertise in the biology and ecology of the coastal environment. Areas of interest broadly include the interactions of marine organisms with one another and with their physical and chemical environment. Research topics may include the ecology of benthic and pelagic organisms, ecological processes and regulatory mechanisms, population recruitment and connectivity, response to habitat changes and disruptions, bio(geo)chemistry and genomic adaptation. Research programs that integrate observational, experimental and modeling approaches in the field and laboratory are encouraged.

Geological, Geophysical and Geochemical Studies of Continental Margins #10-441:
We seek outstanding candidates from throughout the Earth sciences with complementary skills targeting continental margin development, evolution, and physical/chemical exchange across a variety of spatial and temporal scales. An emphasis on the marine portion of the system is envisioned, but candidates investigating onshore aspects of margin processes that cross the shoreline will also be considered. We encourage applicants whose research bears on key processes at margins, including geochemistry, geophysics, dynamics, seismology, tectonics, structural geology, geodesy, petrology, geochemical exchange (high- and low-temperature), geomorphology, glaciology, and ice-margin studies.

All successful candidates will be expected to teach classes, supervise research at both the graduate and undergraduate level and contribute to leadership on issues of equity and diversity. The positions require a PhD degree and a competitive record of publication, as well as evidence of the ability to conduct and fund an active research program consistent with the opportunity to have done so at this career level. The successful candidates will also have demonstrated the highest standards of scholarship and professional activity, or for junior scholars to have the potential thereof. The department is interested in candidates who have demonstrated commitment to excellence by providing leadership in teaching, research or service towards building an equitable and diverse scholarly environment.

Our preference is for hiring at the level of assistant professor for all positions, however appointments at all ranks may also be considered. Salary will depend on the experience of the successful applicant and will be based on the UCSD pay scales. Completed applications received by October 15, 2012 will be assured of consideration. Each position requires separate submission. All applications and related materials must be submitted electronically via Academic Personnel On-Line RECRUIT at: (https://apol-recruit.ucsd.edu). Emailed applications will not be accepted. Applicants should provide a current CV, cover letter including descriptions of their teaching experience, research interests, state if authorized to work in the U.S., and the names of three potential referees, along with their complete institution address, email address, phone and fax numbers, and a separate personal statement summarizing past or potential contributions to diversity (see http://facultyequity.ucsd.edu/Faculty-Applicant-C2D-Info.asp and http://sio.ucsd.edu/Diversity/ for further information). Questions about submission of applications may be addressed to Leslie Costi, (lcosti@ucsd.edu).

UCSD is an Affirmative Action/Equal Opportunity Employer with a strong institutional commitment to excellence through diversity.

NW227624R

Faculty Positions in Chemical Engineering
at the Ecole polytechnique fédérale de Lausanne (EPFL)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

As part of a major initiative, the Institute of Chemical Sciences and Engineering (ISIC) at EPFL invites applications for several faculty appointments in Chemical Engineering. Exceptional applicants with expertise in energy-related topics, including solar energy conversion, chemical, electrochemical and biochemical energy conversion, carbon capture and utilization, are especially encouraged to apply. Appointments at the Assistant Professor level (tenure track) are envisioned, but senior faculty levels (Associate/Full) may also be considered.

A PhD in Chemical Engineering or a related field and an excellent track record of innovative research and leadership are generally required. The successful candidate will be expected to establish and direct a vigorous, independent research program and be committed to excellence in teaching at both the undergraduate and graduate levels.

Applications including cover letter, curriculum vitae, publications list, concise statements of research and teaching interests as well as the names and addresses (including email) of five references should be submitted in electronic format via the website **http://sbpositions.epfl.ch/applications/** by **October 15, 2012.**

The EPFL is consistently evaluated as one of the leading universities in science and engineering in Europe. We offer internationally competitive salaries, start-up resources and benefits. The EPFL aims for a strong presence of women amongst its faculty, and qualified female candidates are strongly encouraged to apply. More information about EPFL and the Institute of Chemical Sciences and Engineering can be found at: **http://www.epfl.ch/** and **http://isic.epfl.ch**. For additional information about this call for applications, please contact the director of the institute, Prof. Paul Dyson **(paul.dyson@epfl.ch)**. W227433R

Change your environment. Find jobs where you'll make a difference

naturejobs

nature REVIEWS

Locum Assistant Editor (Maternity cover)

Nature Publishing Group (NPG) is a division of Macmillan Publishers Ltd, dedicated to serving the academic, professional scientific and medical communities.

Nature Reviews, Nature Publishing Group's noted life science series, includes eight clinical titles that publish high-quality, authoritative review and opinion articles http://www.nature.com/reviews/index.html. We now have an exciting opportunity for a locum Assistant Editor to join our clinical *Nature Reviews* teams. This is a maternity cover position; the initial term of the contract will be full-time for six months. The successful applicant will work on two of the eight titles – *Nature Reviews Rheumatology* and *Nature Reviews Cardiology*.

Our Assistant Editors make an essential contribution to our high-quality content by editing accepted articles for sense, clarity, factual accuracy, and structure, as well as to ensure consistency in style, grammar, punctuation and spelling. Proof-reading and writing content for journal sections are additional key aspects of the role.

Preference will be given to candidates with a good biomedical postgraduate degree and relevant editing experience on clinical or scientific publications. Subject area knowledge for the journals would be useful, but it is not essential. The role is based in our London office.

All applicants must have the right to live and work in the United Kingdom **before they apply**, in order to be considered for the vacancy.

Nature Reviews Rheumatology covers the entire field of rheumatology, from rheumatoid arthritis to bone disease, and provides readers with an up-to-date view of the latest research relating to pathogenesis, diagnosis and treatment.

Nature Reviews Cardiology covers the pathology, diagnosis, treatment and prevention of all types of cardiovascular disease. Basic, translational and clinical aspects of disease process are considered, including discussions of genetics, mechanisms of disease, biomarkers, epidemiology, screening, imaging, intervention and pharmacotherapies.

Closing Date: Monday 20th of August 2012

To apply please send a recent CV and covering letter (merged into a single document) stating your notice period and salary expectation to vacancy **NPG/083/12** on http://jobs.macmillan.com/.

nature publishing group npg

IN227556R

Not paid what you're worth?

Big irritant.

Tactics to improve your salary.

naturejobs

Harvard University
Faculty of Arts and Sciences
Department of Organismic and Evolutionary Biology

TENURE-TRACK OR TENURED PROFESSOR

Position description: The Department of Organismic and Evolutionary Biology seeks to appoint a tenure-track or tenured professor in the field of vertebrate paleontology, emphasizing an evolutionary perspective. We seek an outstanding scientist who will establish an innovative research program and teach both undergraduate and graduate students. We are especially interested in individuals who conduct rigorous, field- and/or laboratory-based analyses of general problems in paleontology and evolution of vertebrate animals, and who employ morphological, functional, developmental, molecular and/or phylogenetic approaches. The successful candidate will receive a curatorial appointment in the Museum of Comparative Zoology (MCZ) with oversight responsibilities for the museum's vertebrate paleontology collection. The appointee will teach and advise at the undergraduate and graduate levels.

Basic qualifications: Ph.D. required by expected start date.

Additional qualifications: Demonstrated excellence in teaching and research is desired. Candidates for a tenured appointment should also evince intellectual leadership and impact on the field and potential for significant contributions to the department, University, and wider scholarly community.

Special instructions: Applications should include a cover letter, a curriculum vitae, statements of research and teaching interests, and representative publications. Review of applications will begin on October 1, 2012:

- Tenured faculty should submit these materials at http://academicpositions.harvard.edu/postings/4228
- Tenure-track applicants should submit these materials, along with the names and contact information of three to five references at http://academicpositions.harvard.edu/postings/4227

Letters of nomination from third parties are also welcome.

Harvard is an Equal Opportunity/Affirmative Action employer. Applications from women and minorities are strongly encouraged.

Contact information: Further information about OEB and MCZ are available at http://www.oeb.harvard.edu and http://www.mcz.harvard.edu.

Contact email: fac-search@oeb.harvard.edu

NW227614R

GEORGETOWN UNIVERSITY

Senior Faculty Position

Theoretical or Computational Soft Condensed Matter Science

Georgetown University invites applications for a senior faculty position (Associate or Full Professor) in the field of computational and/or theoretical soft condensed matter. Applicants should have a broad, internationally recognized record of research accomplishments studying the properties of materials such as complex fluids, gels, emulsions, biomaterials, or active matter, and demonstrated excellence in teaching at the undergraduate and graduate level.

The new hire will join the newly founded Institute for Soft Matter Synthesis and Metrology (http://softmatter.georgetown.edu/), a joint effort of the Departments of Physics and Chemistry. The mission of the Institute is to support research and education in the science of soft matter at Georgetown and in the broader community. Georgetown's commitment to the Institute includes at least two additional hires in soft matter in the near future. It is anticipated that the hire will actively participate in the growth of the Institute and collaborate with current Institute members.

Applicants should forward a CV, a concise statement of research and teaching plans, and contact information for at least four references via e-mail to softmatter@georgetown.edu or to Professor Jeff Urbach, Director, ISMSM, Department of Physics, Georgetown University, Washington, DC 20057. Georgetown University is an Affirmative Action/Equal Opportunity Employer that is committed to diversity in the workplace. Review of applications will begin September 15, 2012 and will continue until the position is filled.

NW227344R

MACMILLAN
SCIENCE COMMUNICATION

Assistant Editor – Macmillan Science Communication

Macmillan Science Communication (MSC) seeks an assistant editor for a range of high profile scientific communication projects. You will be working within an expert team of scientific communications professionals on custom web and print publications produced on behalf of top scientific institutions in the Asia-Pacific region.

The job involves

- Ensuring that the English language content of articles is of high quality, readability and accuracy.
- Copy editing to house style of written content for MSC projects. This includes creating titles, standfirsts, image captions and other short pieces of text as required.
- Uploading content for web production using a content management system (CMS) and web proofing.
- Placing of content for print production using Adobe Indesign and print proofing.
- Ensuring that content is delivered and published to schedule.

The successful candidate will have a sound scientific background with at least a bachelor degree in science, or in a journalism or English-related discipline but with demonstrated fluency in science. The candidate will speak of English to a native level of fluency and will have demonstrable aptitude in scientific and/or technical copy editing. The candidate will receive full training in house styles and procedures. Outstanding written and oral communication skills and the ability to work to tight and changeable deadlines are required. Experience of using content management systems and Adobe Indesign would be an advantage.

Macmillan Science Communication (MSC) is the scientific custom publishing division of Macmillan Publishers Ltd. Macmillan Publishers Ltd. is the owner of Nature Publishing Group (NPG), publisher of the international science journal Nature, Scientific American, and a family of Nature-branded research journals. MSC has exclusive access to the editorial and production expertise of NPG (including Scientific American) in print, online and audio-visual media and thus is able to design and create content with the highest editorial and production standards.

The position will be based at the Asia-Pacific regional headquarters of Macmillan Science Communication in Tokyo, Japan.

Applicants should send a CV, and a brief cover letter explaining their interest in, and suitability for, the post and examples of any previous written work which may be of relevance to:

Matthew Salter, Macmillan Science Communication (Asia-Pacific), Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843 Japan or by email to: jobs@mscasiapacific.com

The closing date for applications is September 7th 2012

IN227429R

ICRISAT
Science with a human face

International Crops Research Institute for the Semi-Arid Tropics

invites applications for **Principal Scientist-Applied Cereals Genomics,** to be based at its Headquarters, near Hyderabad, Andhra Pradesh, India.

Responsibilities: Lead the development and application of genomics for use in ICRISAT's global dryland cereal crop improvement research programs. Conduct relevant research in the development and use of genomic tools for gene identification, genome mapping and marker-assisted breeding in one or more of ICRISAT's dryland cereal crops. Develop linkages and collaborative research projects with advanced research organizations, national agricultural research systems, universities and the private sector. Undertake capacity building initiatives for stakeholders including policy makers to enhance the impact of genomics in rain-fed agriculture. Playing an active role in resource mobilization for ICRISAT.

Requirements: PhD degree in Molecular Biology, Biotechnology, Plant Biology, Plant Breeding, Genetics or other related field. At least 10 years of post-doctoral research experience in plant breeding, molecular genetics, genomics, molecular biology or marker-assisted breeding. A record of achievement with academic publications, development and management of research grants, and capacity building initiatives. Fluency in spoken and written English is essential, and knowledge of French will be an advantage.

Please visit *www.icrisat.org/icrisat-careers.htm for a detailed position announcement and how to send your application.* This is an internationally recruited staff position with competitive pay and benefits payable in US dollars. **The closing date to receive applications is 31 August 2012.**

We strongly encourage women to apply.

RW227555R

Washington University in St. Louis

Tenure Track Assistant Professor Developmental Biologist

The Department of Developmental Biology at Washington University School of Medicine is undergoing a major expansion and invites applications at the level of assistant professor on the tenure track. We are seeking outstanding colleagues with an interest in any area of developmental biology, including the genetic and epigenetic mechanisms of mammalian embryogenesis, cell fate specification and reprogramming, tissue engineering, and the roles of membrane channels and receptors in development and differentiation. Faculty in the Department of Developmental Biology employ a broad range of cell culture systems and model organisms, including C. elegans, D. melanogaster, Xenopus, zebrafish and mouse. (Please visit our website at http://devbio.wustl.edu/index.htm)

Review of applications will start October 15, 2012. Interested applicants should email a single PDF file of their curriculum vitae, summary of their research accomplishments and plans, and should also arrange to have three letters of recommendation emailed to:

Search Committee
(c/o Lilianna Solnica-Krezel, Ph.D.)
Department of Developmental Biology
Washington University School of Medicine
660 South Euclid Avenue
Campus Box 8103
St. Louis, MO 63110
Email: devbiosearch@wustl.edu

Washington University is an Equal Opportunity Employer AA/EOE M/F/D/V.

RW227628R

FRED HUTCHINSON CANCER RESEARCH CENTER
A LIFE OF SCIENCE

The Division of Basic Sciences invites applications at the **junior faculty member** level. We are seeking outstanding scientists conducting cutting-edge research into fundamental biological questions. Current faculty members investigate diverse areas of genetic, molecular, cellular, metabolic, developmental, evolutionary, structural and computational biology and utilize a broad range of approaches and experimental systems. We support our junior faculty by providing a collegial environment and a wealth of technical resources. The Fred Hutchinson Cancer Research Center is located in a modern campus by Lake Union in Seattle, Washington, and is close to other non-profit research institutes and the University of Washington. It has an outstanding graduate program and state of the art facilities and research resources. Further information about the Center is available at **http://www.fhcrc.org/basic**.

Candidates should visit **http://www.fhcrc.org/basicfacsearch** for application instructions.

Application deadline:
November 9, 2012

The Fred Hutchinson Cancer Research Center is an equal opportunity employer, committed to workforce diversity.

NW227646R

אור ותמים
LUX ET VERITAS

Yale University

Faculty Position in Plant Ecology

The Department of Ecology and Evolutionary Biology at Yale University invites applications for a tenure-track assistant professor position in plant ecology, including interactions between plants and other organisms and biogeochemical cycles. We seek applicants that use creative approaches to address fundamental questions in ecology. We are particularly interested in applicants that have or would develop a strong field research program. A record of outstanding achievement and a promising research program are more important than the specific research area.

Interested candidates should submit a CV, three relevant reprints or manuscripts, brief research and teaching statements, and names and addresses of three potential evaluators online at https://academicjobsonline.org/ajo/Yale/EEB. The search will remain open until the position is filled. The review of applications will begin 1 September 2012.

Yale University is an Equal Opportunity/Affirmitive Action Employer. Yale values diversity among its students, staff and faculty and strongly welcomes applications from women and underrepresented minorities.

NW227585RL

nature jobs

Scientists stay informed of the latest:

- **Hot jobs**
- **Career news**
- **Announcements**

Sign-up to our **FREE** jobseeker newsletter:
naturejobs.com/newsletter

REPROGRAPHICS: CC Media Group, London EC1V 8AR UK and **The Charlesworth Group,** Wakefield, UK.
PRINTED BY: William Gibbons & Sons Ltd, UK; **Publishers Press,** Lebanon Junction, Ky, USA and **Obun Printing Co. Inc,** Tokyo, Japan

a nature conference

2012 *Nature* / Institute for Genomic Medicine Symposium

Building upon the outstanding success of the Inaugural and 2011 Symposia at the Institute for Genomic Medicine at University of California, San Diego, this year's symposium, in partnership with *Nature*, is aimed at the international scientific community, with the goal of fostering and facilitating communication and collaborative initiatives in medical genomics.

The following topics will be covered at the conference: Cancer Genomics, Gene & Cell Therapies, Complex Human Traits, Personalized Genomics for Implementation of Genetic Information.

November 4-5, 2012
UCSD School of Medicine Campus
La Jolla, CA, USA

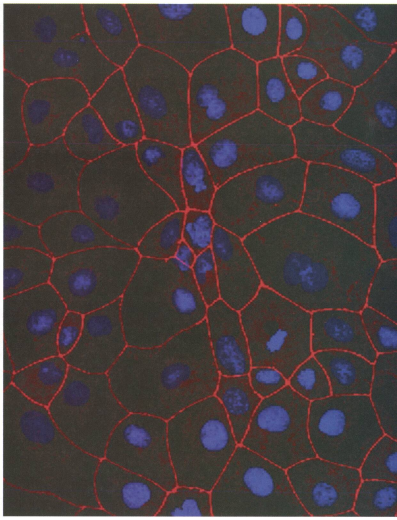
KEYNOTE SPEAKER:
Eric Green (National Human Genome Research Institute, NIH, USA)

CONFIRMED SPEAKERS:

Robin Ali
Ines Barosso
Jean Bennett
George Church
Mark Daly
Richard Gilbertson
Joseph Gleeson
Joe Gray
David Haussler
Vanessa Hayes
Jonathan Sebat
Mike Snyder
Mike Stratton
Doug Turnbull

ORGANIZERS:
Kelly Frazer (University of California, San Diego, USA)
Xiang-dong Fu (University of California, San Diego, USA)
Magdalena Skipper (*Nature*, UK)
Kang Zhang (University of California, San Diego, USA)
Kun Zhang (University of California, San Diego, USA)

For more information and to register visit:
www.nature.com/natureconferences/nigm2012



IGM Institute for Genomic Medicine

UC San Diego Health Sciences

nature

IN226557E

SKLID 传染病诊治国家重点实验室
State Key Laboratory for Diagnosis and Treatment of Infectious Disease

MACMILLAN SCIENCE COMMUNICATION

Scientific Writing and Publishing Workshops

The State Key Laboratory for Diagnosis and Treatment of Infectious Diseases of Zhejiang University,

and

Macmillan Science Communication (MSC), an exclusive partner of Nature Publishing Group,

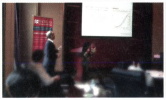
are holding two Scientific Writing and Publishing Workshops in Hangzhou, 12th – 15th September 2012.

- The workshops will train researchers of all levels to prepare and publish high quality research papers for publication in international, peer-reviewed journals. We will discuss in-depth the issues that researchers face today. There will be one 2-day intensive workshop for experienced researchers and one 1-day workshop for researchers and students with less publishing experience.
- Training will be led by *Nature* Executive Editor, Dr. Nick Campbell, and *Nature Reviews Genetics* Senior Editor, Dr. Mary Muers.
- Following the workshops, Dr. Peter Gorsuch, Science Editor at MSC and former *Nature* editor, will provide independent feedback on selected, unpublished papers written by workshop attendees in individual consultations.

For more information about Zhejiang University:
www.zju.edu.cn/english

For more information about the MSC workshops:
www.msceditting.com

For more information about The State Key Laboratory for Diagnosis and Treatment of Infectious Diseases:
www.sklid.com



This workshop is brought to you by Macmillan Science Communication, exclusive partner of Nature Publishing Group, publisher of *Nature*.

nature publishing group npg

IN2272167

Royal Society scientific meetings

September 2012 – July 2013

Registered Charity No. 207043

Scientific discussion meetings at the Royal Society, London

Magnetoelectric phenomena and devices
24 – 25 September 2012
Organised by Dr Neil Mathur and Professor James Scott FRS

Regulation from a distance: long-range control of gene expression in development and disease
22 – 23 October 2012
Organised by Professor Wendy Bickmore and Professor Veronica van Heyningen FMedSci FRS

Energy transduction and genome function – an evolutionary synthesis
12 – 13 November
Organised by Dr Nick Lane, Professor John F Allen, Professor John Raven FRS and Professor William Martin

Achieving food and environmental security – new approaches to close the gap
3 – 4 December
Organised by Professor Guy Poppy, Professor Paul Jepson, Professor John Pickett CBE FRS and Dr Michael Birkett

Taking x-ray phase contrast imaging into mainstream applications
11 – 12 February
Organised by Dr Alessandro Olivo and Professor Ian K Robinson

Characterising exoplanets: detection, formation, interiors, atmospheres and habitability
11 – 12 March
Organised by Professor Steve Miller, Professor Jonathan Tennyson FRS, Professor Athena Coustenis and Professor Peter Read

Cellular polarity: from mechanisms to disease
15 – 16 April
Organised by Dr Rafael Edgardo Carazo Salas, Dr Attila Csikasz-Nagy and Dr Masamitsu Sato

e-futures: beyond Moore's Law
13 – 14 May
Organised by Professor David Cumming, Professor Steve Furber CBE FREng FRS and Professor Douglas Paul

Language in developmental and acquired disorders: converging evidence for models of language representation in the brain
10 – 11 June
Organised by Professor Dorothy Bishop FMedSci FBA, Professor Karalyn Patterson FMedSci FBA and Professor Kate Nation

All of the above discussion meetings will be followed by a focused satellite meeting at the Royal Society at Chicheley Hall, home of the Kavli Royal Society International Centre. See our meeting website listing for details.

Theo Murphy meetings at the Royal Society at Chicheley Hall, home of the Kavli Royal Society International Centre, Buckinghamshire

Complex patterns in wave functions – drums, graphs and disorder
5 – 7 September
Organised by Dr Sven Gnutzmann and Professor Uzy Smilansky

Handling uncertainty in weather and climate prediction, with application to health, agronomy, hydrology, energy and economics
4 – 5 October
Organised by Professor Tim Palmer FRS

UK-China workshop on the chemistry and physics of functional materials
28 – 29 January
Organised by Professor Henning Sirringhaus FRS, Professor Anthony Cheetham FRS, Professor Wenping Hu and Professor Deqing Zhang

Storage and indexing of massive data
7 – 8 February
Organised by Professor Costas Iliopoulos, Dr Simon Puglisi and Professor Maxime Crochmore

Space in the brain: cells, circuits, codes and cognition
1 – 3 May
Organised by Dr Tom Hartley, Professor John O'Keefe FRS, Professor Neil Burgess and Dr Colin Lever

New models and observations of the Southern Ocean, its role in global climate and the carbon cycle
16 – 17 July
Organised by Professor Andrew Watson FRS, Professor John Marshall and Dr Mike Meredith

These meetings are free to attend, but pre-registration is essential.

For more details visit **royalsociety.org/events**

THE ROYAL SOCIETY

U227266EL

TIGE IS THE MAN

An old-fashioned night out.

BY SUE LANGE

Tige is the man and he knows it. Mr Personality. Sharp-tie-wearing, punchline-busting, girl-sweet-talking. Everybody loves him; he puts up such a good front. And his parties? Sublime.

Tonight he's gone all out, giving the cats something they've never tasted. A trip into the past when things were dicey and anything could happen. Not like today where you know what's going to happen even before it does.

The guests arrived at Tige's in twos, couples holding tight as if in love. As if someday they'd reveal themselves to their one and only. Expose their naked feelings to their significant other late at night, maybe, when the curtains were closed and the children were in bed.

He watched them and wondered and laughed to himself. He wasn't even drunk yet, but he laughed, because he knew that before the night was through all would be revealed.

He'd spiked the rum punch with floating hothouse gardenias. He'd padded each ash-tray with green marijuana spliffs. His house delivery system was set to a groovalicious pop rock. The beat, key and bassline changed with each new arrival at the door. As the guests entered, they let their guard down for half a mo' so that the house could read their minds and respond with the appropriate musical vibe. The evening was electric.

At ten, the games began.

"Let us hark back to the age of chivalry," Tige said from his post in front of the defunct fireplace in the wall of his condo. He had a malicious grin on his face when he said: "When goils were goils, and men were men."

Everyone knew the bit of mass culture he was referring to. They knew it all by heart. And if they didn't, they could check it remotely, instantaneously.

Tige continued. "I'm talking about the late twentieth, ladies and gents. Back when people had open minds. Privacy was their birthright. It was easy to be a man or a goil back then."

They laughed again. They didn't really know why. But the words sounded funny so they laughed.

"They had party games like we do, but they were different," Tige said. "They played Charades and Pin the Tail on the Donkey instead of our games like My 15 Minutes or Pass the Vid. There was one such oddity called Truth or Dare, and that's the one we'll hark back to tonight. It's all about daring to tell the truth. And daring to believe it. Remember boys and goils, it was a leap of faith back then. Back then, no one could check. Nowadays…" Tige moved his eyebrows up and down as if he was peeking under somebody's skirt.

"Und so!" he shouted with finger pointing to the ceiling. "Tonight we check!"

He waved his hands in the air, beckoning to the house and its mind-readingware installed in the walls. "The very latest in Sony Hive-Mind tech," he said, as if that was a statement, information they needed because they couldn't have guessed that Tige wouldn't be caught dead without the latest in anything.

"Tonight we make history," Tige declared. "For the first time in humanity's long flight from the trees, we dare to be truthful."

And with one final flourish, he passed his hand in front of his face and let down his guard.

The crowd of half-drunk, half-high, neo-hip-cons gasped. One young girl tittered as she saw something she'd never seen before: a big long thought. Tige remained steadfast even as his face turned a bit rose. He grew bolder and took in the faces of his guests, passing judgments as he scanned. They could see what he thought of them, who he liked and who he'd screwed and who he'd rather see dead.

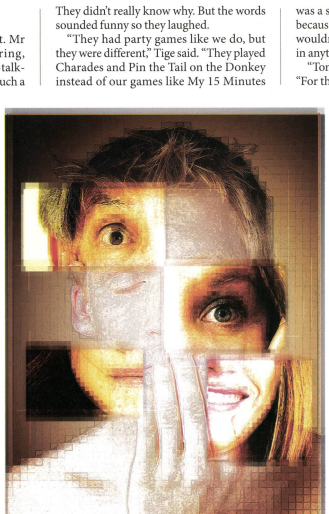
They laughed and cried, guffawed and swore; he made some enemies in a quick five minutes. But his fears slowly left him and his conscience ran clear. A smile spread across his face. He felt fine.

"Come on in, my friends," he said. "It's great! A little draughty in the old noggin maybe, but in general I feel fine."

The crowd remained silent and thoughtful for a moment only and then they, not wanting to be left behind at the avant-garde event of the year, likewise passed their hands in front of their faces with a flourish. They let slip their shields and revealed what lay behind.

The house extracted the contents of their minds and published them in Tige's room. They pointed at each other and giggled with hands over their mouths as if trying to hold back thoughts that burbled out. Some turned scarlet; one man ran to the bathroom and cried. The more they drank, though, the more they howled. A good time was had by all.

At five in the morning, when dawn threatened to break, they sobered up and replaced the shields. They cleared their throats, adjusted their caps, straightened their skirts. One by one, they left alone. ■



JANSKY

➲ NATURE.COM
Follow Futures on Facebook at:
go.nature.com/mfoodm

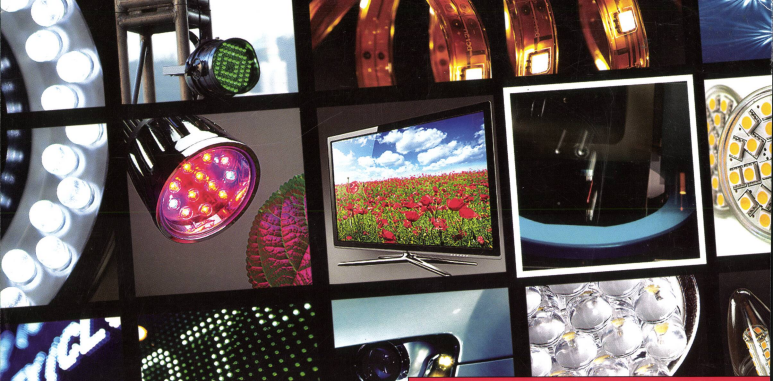
Sue Lange *is a writer living in Pennsylvania. Her parties are not to be missed.*

nature podcast

science... wherever you are.

nature.com/nature/podcast

nature publishing gro



STEALTH DICED LEDs

BRIGHTER WORLD

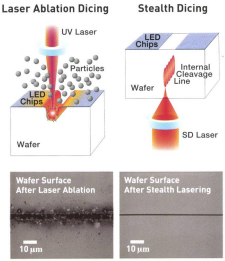
Hamamatsu Photonics "Stealth Dicing" technology helps produce better, brighter, more affordable LEDs to enlighten our world.

Cutting an LED semiconductor wafer into individual chips with a laser, as manufacturers generally do, produces heat and destroys some of the reflective surface layers that help to channel light. This results in dimmer LEDs.

Hamamatsu Photonics solved these problems by reinventing a decade-old technology named Stealth Dicing. With Stealth Dicing, instead of cutting across the surface of the wafer, the laser is focused beneath the surface, creating a precise, "hidden" rift. The chips then break apart neatly.

No water is needed, saving a precious resource. Also, the width of the rift is nearly zero, sparing the reflective surface and thus making brighter LEDs for less money.

These improved efficiencies open the door for widespread use of these energy saving lights in general illumination and paves the way for other applications of Stealth Dicing. For example, stealth-diced silicon carbide chips could be used in next generation electronic control systems in electric cars, photovoltaics and smart grids.



Stealth Dicing eliminates the molten debris of laser ablation dicing, so it requires no water clean-up.

Hamamatsu Photonics' expertise with light has revealed an elegant solution to a complex problem.

HAMAMATSU

PHOTON IS OUR BUSINESS

For more information, visit our site at: http://jp.hamamatsu.com/en/rd/publication

HAMAMATSU PHOTONICS K. K.

325-6, Sunayama-cho, Naka-ku, Hamamatsu City, 430-8587 Japan Telephone: (81)53-452-2141, Fax: (81)53-456-7889 URL http://www.hamamatsu.com